Vol. 43, Parts 1 and 2 June 1956

QH
301
B615

UNIVERSITY
OF MICHIGAN
JUN 28 1956
SCIENCE
LIBRARY

BIOMETRIKA

FOUNDED BY

W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

MANAGING EDITOR

E. S. PEARSON

ASSOCIATE EDITORS

M. G. KENDALL JOHN WISHART

in consultation with

HARALD CRAMÉR J. B. S. HALDANE
F. N. DAVID H. O. HARTLEY
R. C. GEARY D. G. KENDALL

ISSUED BY

THE BIOMETRIKA OFFICE, UNIVERSITY COLLEGE, LONDON

PRINTED AT THE UNIVERSITY PRESS, CAMBRIDGE

[*Issued* 4 *June* 1956]

A volume of *Biometrika* containing about 500 pages, with plates and tables, will be published annually in two half-yearly issues.

Papers for publication should be sent either to

PROFESSOR E. S. PEARSON

Department of Statistics, University College, London, W.C. 1

or if more convenient to

DR JOHN WISHART

Statistical Laboratory, St Andrew's Hill, Cambridge

or to PROFESSOR M. G. KENDALL

London School of Economics, Houghton Street, London, W.C. 2

It is a condition of publication in *Biometrika* that the paper shall not already have been issued elsewhere, and will not be reprinted without leave of the Editors.

Contributors receive 25 copies of their papers free. Joint authors 15 copies each. Order forms for separates are sent to authors with proofs of their papers.

The Subscription price, payable in advance, is:

Inland: 45*s.* net per volume Abroad: 54*s.* net per volume

including packing and postage

BACK ISSUES

Volumes 20–42

These may be obtained from the BIOMETRIKA OFFICE, at the following prices:

Vols. 20–37: £5 per volume Vols. 38–42: £3. 3*s.* per volume

packing and postage 2s. per volume

Bound volumes: £1 extra per volume Binding cases: 10*s.* each

Cheques should be made payable to *Biometrika*, crossed "a/c BIOMETRIKA TRUST" and sent to

THE SECRETARY
BIOMETRIKA OFFICE
UNIVERSITY COLLEGE
GOWER STREET, LONDON, W.C. 1

to whom all orders for series, single copies and offprints should be addressed.

All foreign cheques must be drawn in sterling and on a Bank having a London Agency.

Volumes 1–19

Permission for reprinting has been granted to Messrs Wm. Dawson & Sons Ltd. Volumes 1–11 are now ready for distribution and Volumes 12–19 are in course of preparation. Would librarians and others wishing to have copies please place their orders now with:

WM. DAWSON & SONS LTD.,
4 DUKE STREET, MANCHESTER SQUARE,
LONDON, W. 1

VOLUME 43, PARTS 1 AND 2 JUNE 1956

739
8813A

STUDIES IN THE HISTORY OF PROBABILITY AND STATISTICS

II. THE BEGINNINGS OF A PROBABILITY CALCULUS

BY M. G. KENDALL

Research Techniques Unit, London School of Economics and Political Science

1. The first article in this series by Dr F. N. David (1955) has reviewed the development of dicing and gaming up to the time of Fermat and Pascal, who are popularly but erroneously supposed to have founded the calculus of probability. In this paper I shall try to trace the evolution of the idea of a probability calculus with especial reference to what Dr David calls the tantalizing period prior to A.D. 1600.

2. During the dark ages gaming was prevalent throughout Europe. At some unknown point of time dice finally ousted tali as instruments of play; and since cards were not introduced until about A.D. 1350 gaming must have been conducted mainly with dice for nearly a thousand years. Efforts on the part of Church and State to control the evils associated with it were as ineffectual then as they are today, and nothing is more indicative of the persistence of gambling than the continual attempts made to prevent it. The sermon of St Cyprian of Carthage *De Aleatoribus* (*c.* A.D. 240) was echoed twelve hundred years later in the more famous sermon of St Bernardino of Siena *Contra Alearum Ludos* of A.D. 1423. The gambling of the Germans referred to by Tacitus may perhaps have become more moderate but was equally prevalent in the thirteenth century when we find Friedrich II (1232) issuing a law *de aleatoribus* and Louis IX (1255) forbidding not only the play but even the manufacture of dice. A long series of edicts prohibiting the clergy from gaming (e.g. by Otto der Grosse, A.D. 952, the Councils of Trèves in A.D. 1227 and 1235, the Council of Worcester in A.D. 1240) are themselves eloquent of the failure on the part of the authorities to repress the evil.

3. We must remember, however, that all these banns and prohibitions were not really directed against games of chance as such but against the vices which accompanied them. There seems to have been nothing impious in creating a chance event or in using it for purposes of amusement. The Church was much more concerned about the drinking and swearing which accompanied gaming; and the State was more concerned about the idleness, thriftlessness and crime which were so often found among gamblers. Chaucer's Pardoner puts the official view of his day by giving an example of the blasphemy which usually accompanied a gambling game (almost certainly of hazard)*

> By Goddës precious heart and by his nails
> And by the blood of Christ that is in Hayles
> Seven is my chance, and thine is cinq and trey.
> By Goddës armës, if thou falsely play
> This dagger shall throughout thine hertë go!—
> This fruit cometh of the bitchëd bonës two:
> Forswearing, irë, falseness, homicide.

Even chess, the most innocent of all games, was classed among the major vices, at least for officials. The interdict of Louis IX referred to above says: 'They shall abstain...from

* Here and elsewhere I have modernized the spelling of the English quotations to some extent where the metre permits.

dice and chess, from fornication and frequenting taverns. Gaming-houses and the manufacture of dice are prohibited throughout the realm.'*

4. San Bernardino enumerates at length fifteen 'malignitates impiisimi ludi' but they are all moral evils (love of gain, idleness, corruption of youth, etc.) with the exception of blasphemy and contempt of the prohibitions of the Church. One feels that if there had been anything to say about the impiety of eliciting chance events for innocent entertainment the saint would certainly have said it. The general attitude of his time seems to have been one of toleration of the actual play but stern opposition to its associated vices. There is some positive evidence to the same effect. About A.D. 960 a certain bishop Wibold of Cambray invented a clerical version of dice to which I shall refer later; this sagacious realist evidently recognized the impossibility of stamping out the evil and hence attempted to turn it into good. Participants in the third crusade (A.D. 1190) had, in their briefing instructions, a carefully drawn up statement of the extent to which they might gamble; no person below the rank of knight was permitted to play at all for money; knights and the clergy might play but could not lose more than twenty shillings in twenty-four hours. Chaucer, in the Franklin's Tale, refers to the playing of chess and tables (backgammon) with the laudable object of distracting the heartbroken Dorigene. In 1484 Margery writes to John Paston (24 December):

> Please it you to wit that I sent your eldest son to my Lady Morley to have knowledge what sports were used in her house in Christmas next following after the decease of my lord her husband; and she said that there were none disguisings, nor harping, nor luting, nor singing nor none loud disputes; but playing at the tables (backgammon) and chess and cards; such disports she gave her folks leave to play and none other.

5. We may also notice a series of laws, beginning in the reign of Edward III, prohibiting the playing of certain games in order to promote manly sports. An act of Henry VIII added dice and cards to the list of unlawful amusements, although Henry, like many other monarchs, set a very bad example to his subjects. These laws were militaristic in origin. The common people were not to waste their leisure in playing peaceful games like bowls, ninepins, hockey and dice; their duty was to practice archery in readiness for the next war.

6. I recall these facts to establish two points. The first is that playing with dice (and later with cards) continued from Roman times to the Renaissance without interruption and was practised not only among the educated classes but among the middle classes and among the lower classes also. The second is that although the various Governments and the Church discouraged gaming to the point of prohibition, a great deal of play went on either as innocent pastime or, by popular approval, in defiance of the law.

7. One of the exasperating features of the many references to dice-playing between A.D. 1000 and 1500 is that authors invariably assume that their readers are familiar with the games they mention; and hence no rules of play are offered. We are thus very much in the dark about the exact nature of the games which were played. There are two in particular which have a long and interesting history: hazard, the ancestor of the modern American crap game, and primero, the ancestor of poker. It is instructive to consider briefly their line of development.

* 'Abstineant...a ludo etiam cum taxillis vel aleis vel scacis, et a fornicatione et tabernis. Scolas etiam deciorum prohibemus omnino....Fabrica vero deciorum prohibeatur ubique in nostro regno.' Taxillus is a diminutive of talus, but I do not know whether in this context it refers to the talus or the die.

8. The Romans played with four tali (huckelbones) but with only three tesserae (dice). At some early stage versions of dice-playing with only two dice are mentioned; for example, Bishop Eustathius, in a commentary on the Odyssey written in A.D. 1180, refers to games with two dice. The Chaucerian extract quoted above also mentioned two dice. In 1707 Montmort wrote of 'le quinquenove, le jeu de trois dez et le jeu du hazard. Les deux premiers sont les seules jeux de dez qui soient en usage en France, le dernier n'est commun qu'en Angleterre.' Both quinquenove and hazard were played with two dice; all three games are variants of the same idea.

9. Hazard, the game as distinct from its modern meaning of chance in general, was, I believe, brought back to Europe by the Third Crusaders. Godfrey of Bouillon gives a false derivation: 'À Hazait [Hazar] s'en ala ung riche mandement, et l'apiel-on Hazait pour le fait proprement que ly dés fu fais et poins premierment.' There can be little doubt that the word derives from the Arabic *al zhar*, meaning a die.* Wherever it came from, the name and the game must have spread rapidly through Europe. Jean Bodel's play *Le Jeu de Saint Nicolas*, ascribed to the year A.D. 1200, refers to *hazart*. Salimbene (the son of a crusader), writing about 1287, refers to playing 'ad azardum alias ad taxillum'. Dante's *Purgatorio*, written between A.D. 1302 and 1321, refers to *azar*; and Chaucer (about A.D. 1375) uses the word several times. The exact rules of play are not, so far as I am aware, on record. There were doubtless many variants. But the quotation from Chaucer given above suggests that the essential features of the modern game of craps were present at an early date; the addition of the numbers on two dice and the 'chances' of each player are clearly indicated.†

10. It might have been supposed that during the several thousand years of dice playing preceding, say, the year A.D. 1400, some idea of the permanence of statistical ratios and the rudiments of a frequency theory of probability would have appeared. I know of no evidence to suggest that this was so. Up to the fifteenth century we find few traces of a probability calculus and, indeed, little to suggest the emergence of the idea that a calculus of dice-falls was possible. It may be that gamblers had a rough idea of relative frequencies of occurrence—it is hard to see how they could fail to acquire such a thing; and as there is some evidence of the manufacture of false dice from Roman times onwards there was presumably a complementary notion of fair throwing. It may also be that some intelligent man worked out the elements of a theory for himself but guarded his secret on account of its cash value. But I do not really believe this. Other people tried to do the same thing later, but not with permanent success.

11. The earliest work I know which would seem to have mentioned the number of ways in which dice can fall is the game invented by Wibold, referred to above. So far as I am aware no contemporary manuscript has survived but an account of the game (a very obscure one, incidentally) was given by the chronicler Baldericus in the eleventh century, the work being first published in 1615. Wibold enumerated 56 virtues—one corresponding

* Libri (1838–41, vol. 2, p. 188) also gives a false derivation: 'Ce mot vient d'asar, qui en arabe signifie *difficile*', the difficulty in question being that of obtaining two aces or two sixes. Libri undoubtedly got this from the Dante Commentary mentioned in section 14 below.

† The 'chance' at hazard is not a probabilistic one. A player either calls or throws a 'main' (e.g. in one version, any number from 5 to 9 inclusive). He then throws again and may (*a*) 'throw out', in which case the dice pass to his opponent; (*b*) win outright by throwing a certain score; (*c*) throw another score which becomes his 'chance'. He then goes on throwing until either the 'main' or his 'chance' turns up, losing in the first and winning in the second case.

to each of the ways in which three dice can be thrown, irrespective of order. Apparently a monk threw a die three times, or threw three dice, and hence chose a virtue which he was to practice during the next 24 hours. It does not sound much of a game, but perhaps I have misunderstood Baldericus's account. The important point is that the partitional falls of dice were correctly counted. There was no attempt at assessing relative probabilities.

12. The use of dice for the purpose of choosing among a number of possibilities may well be much older than Wibold and certainly continued for long after his time. There exist several medieval poems in English, setting out the interpretations to be placed on the throws of three dice. The best-known is the *Chaunce of the Dyse* which is in rhyme royal; one verse for each of the 56 possible throws of three dice. For example, the throw 6, 5, 3 gives

> Mercury that disposed eloquence
> Unto your birth so highly was incline
> That he gave you great part of science
> Passing all folkës heartës to undermine
> And other matters as well define
> Thus you govern your wordës in best wise
> That heart may think or any tongue suffise.

Another incomplete poem in the Sloane manuscripts also deals with the throws systematically but in a different manner; e.g. for 6, 5, 3,

> Thou that has six, five and three
> Thy desire to thy purpose may brought be
> If desire be to thee y-thyght
> Keep thee from villainy day and night.

Poems like these were doubtless used for elementary fortune-telling—one threw the dice to pick the phrase peculiar to oneself. Those mentioned come from what are probably early fifteenth-century manuscripts and have a certain interest in connexion with divination probabilities. For my present purposes the point to be noticed is that, for purely astragalomantic reasons, the different possible throws were enumerated and known without any reference to gaming or a probabilistic basis.

13. A similar idea is expressed in San Bernardino's sermon of A.D. 1423. The Saint makes a very detailed comparison of the Church of Christ and the church of Satan, represented in this instance by gaming. The Church corresponds to the gaming house; the altar to the gaming table; the sacrificial vessels to the dice box; and so on. In the middle of all this nonsense occurs a passage which does a little to compensate us for having to read it. 'The missal I compare to the die; for in flexibility, permanence and scope it is in no way inferior to the missal of Christ himself; and just as that missal is composed of a single alphabet of twenty-one letters, so in the [game of] dice there are twenty-one throws.'*

The twenty-one possible throws are undoubtedly those with two dice. This number is correct on the interpretation of the indistinguishability of partitions. One cannot help but admire the twisted ingenuity of the comparison, or speculate on what the Saint would have said had he been dealing with the gospel in Greek or Aramaic.

14. The earliest approach to the counting of the number of ways in which three dice can fall (permutations included) appears to occur in a Latin poem *De Vetula*. This remarkable work was regarded as Ovid's for some time and is included among certain medieval

* 'Missale vero taxillum, esse volo: qui quidem, et tractabilior, et durabilior, atque continentia non erit minor, quam sit missale ipsius Christi, cum in eius missali solum alphabetum, hoc est viginti una literae comprehendantur, ac totidem puncta in decio concludantur.'

editions of his poems. It is, however, supposititious and several candidates have been proposed for authorship. The one generally preferred is Richard de Fournival (1200–50), a gifted humanist of the Middle Ages and Chancellor of the cathedral of Amiens. If this is correct the poem was presumably written between A.D. 1220 and 1250. It contains a long passage dealing with sports and games, and with dicing in particular.* It is, perhaps, worth giving in full what is (if genuine) the first known calculation of the number of ways of throwing three dice. The text (taken from an edition published at Wolfenbüttel in 1662) is given in an appendix (pp. 13–14 below).

The relevant passage may be briefly and freely construed as follows:

If all three numbers are alike there are six possibilities; if two are alike and the other different there are 30 cases, because the pair can be chosen in six ways and the other in five; and if all three are different there are 20 ways, because 30 times 4 is 120 but each possibility arises in 6 ways. There are 56 possibilities.

But if all three are alike there is only one way for each number; if two are alike and one different there are three ways; and if all are different there are six ways. The accompanying figure shows the various ways.

[It follows, but is not stated, that the total number of ways is

$$(6 \times 1) + (30 \times 3) + (20 \times 6) = 216.]$$

15. The accompanying figures are shown in Plates 1 & 2, taken from Harleian MS. 5263. The figure referred to above is also given in my Fig. 1, taken from the printed edition published at Wolfenbüttel in 1662. The total of the last column is 108 which, doubled, gives us the total number of ways of throwing three dice. If this is a thirteenth-century product (the manuscript seems to be fourteenth century) it is astonishingly in advance of its time; and some of the phrases have a very modern ring (e.g. 'tria schemata surgunt', three cases arise; 'quemlibet cum dederis, reliqui duo permutant loca', if you fix one, the others permute in two ways).

16. There exists a medieval translation into French of the *De Vetula*, edited and published in 1862 by H. Cocheris, who is mainly responsible for the theory that de Fournival was the author. The translation takes considerable liberties with the original text and is not always easily matched against it. This poem is attributed to the fourteenth century. So far as I can see, the translator seems to have failed to understand the main point. He merely enumerates the 16 possible scores with three dice and points out that some of them occur more often than others. The essential step in the *De Vetula* has been lost.

17. In the sixth canto of the *Purgatorio* Dante mentions the game of hazard:

Quando si parte il giuoco della zara
Colui che perde si riman dolente
Ripetendo le volte e tristo impara:

(When a game of hazard breaks up the loser remains behind mournfully recalling the throws and learning by sad experience.)

A commentary on this passage published in 1477 says

Concerning these throws it is to be observed that the dice are square and every face turns up, so that a number which can appear in more ways [*sc.* as the sum of points on three dice] must occur more frequently, as in the following example: with three dice, three is the smallest number which can be thrown, and that only when three aces turn up; four can only happen in one way, namely as two and two aces.

* I am not competent to express an opinion about the attribution of the *De Vetula* to de Fournival, but I have considerable doubts on the point if the later printed versions correctly record what the author wrote about dice. Some of the critical passages may, however, be interpolations by later hands.

At this point the author seems to be on the verge of the usual fallacy that a three will occur equally as frequently as a four. But (more by luck than knowledge, in my view) he veers off this point and proceeds: 'and so, as these numbers can only happen in one way at each throw, in order to avoid tedium and too long a wait, they are not reckoned in the game and are called hazards. And so for 17 and 18. . . .The numbers in between can happen in more ways; the number which can happen in most ways is said to be the best throw of the set.'

Quinquaginta modis & ſex diverſificantur
In punctaturis, punctaturæque ducentis
Atque bis octo cadendi ſchematibus, quibus inter
Compoſitos numeros, quibus eſt luſoribus uſus,
Diviſis, prout inter eos ſunt diſtribuendi,
Plenè cognosces, quantæ virtutis eorum
Quilibet eſſe poteſt, ſeu quantæ debilitatis:
Quod ſubſcripta poteſt tibi declarare figura.

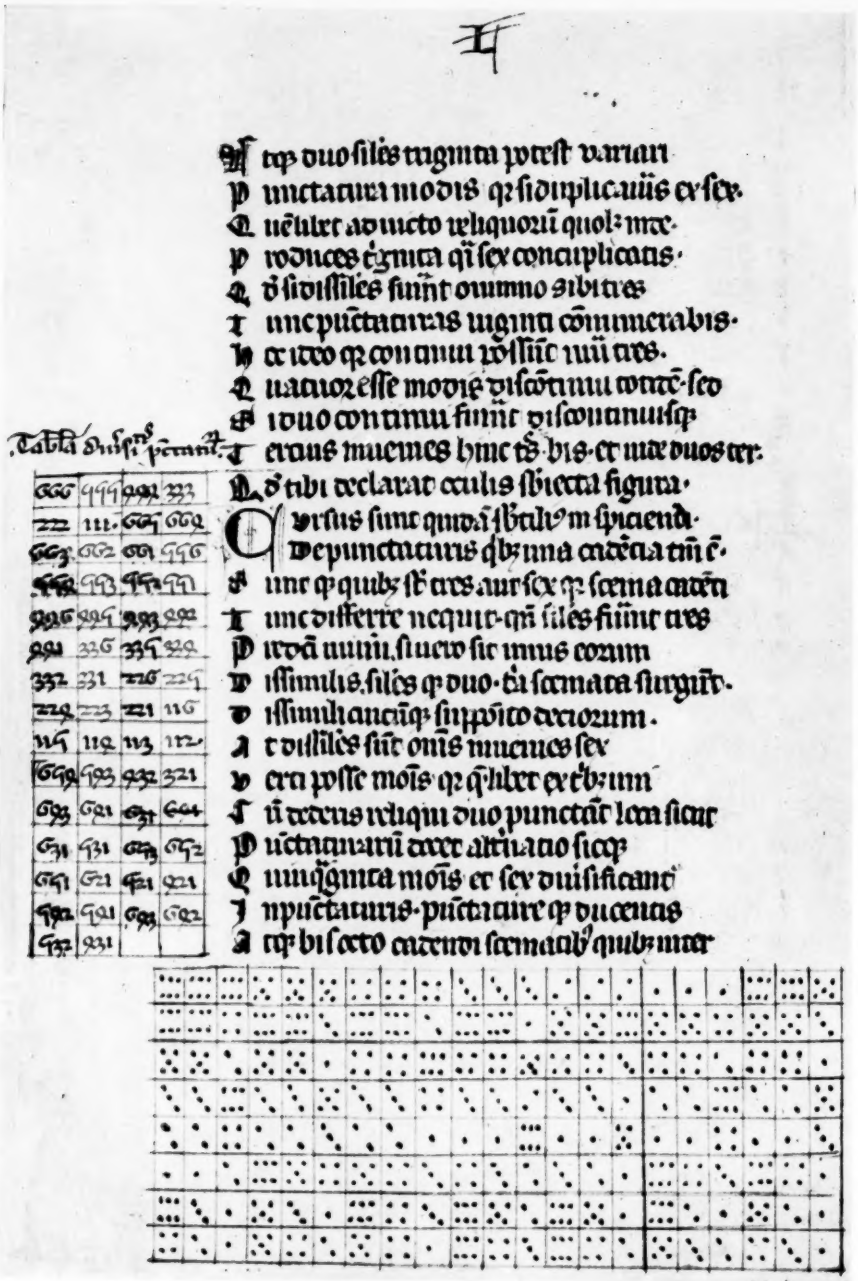
Tabula III.

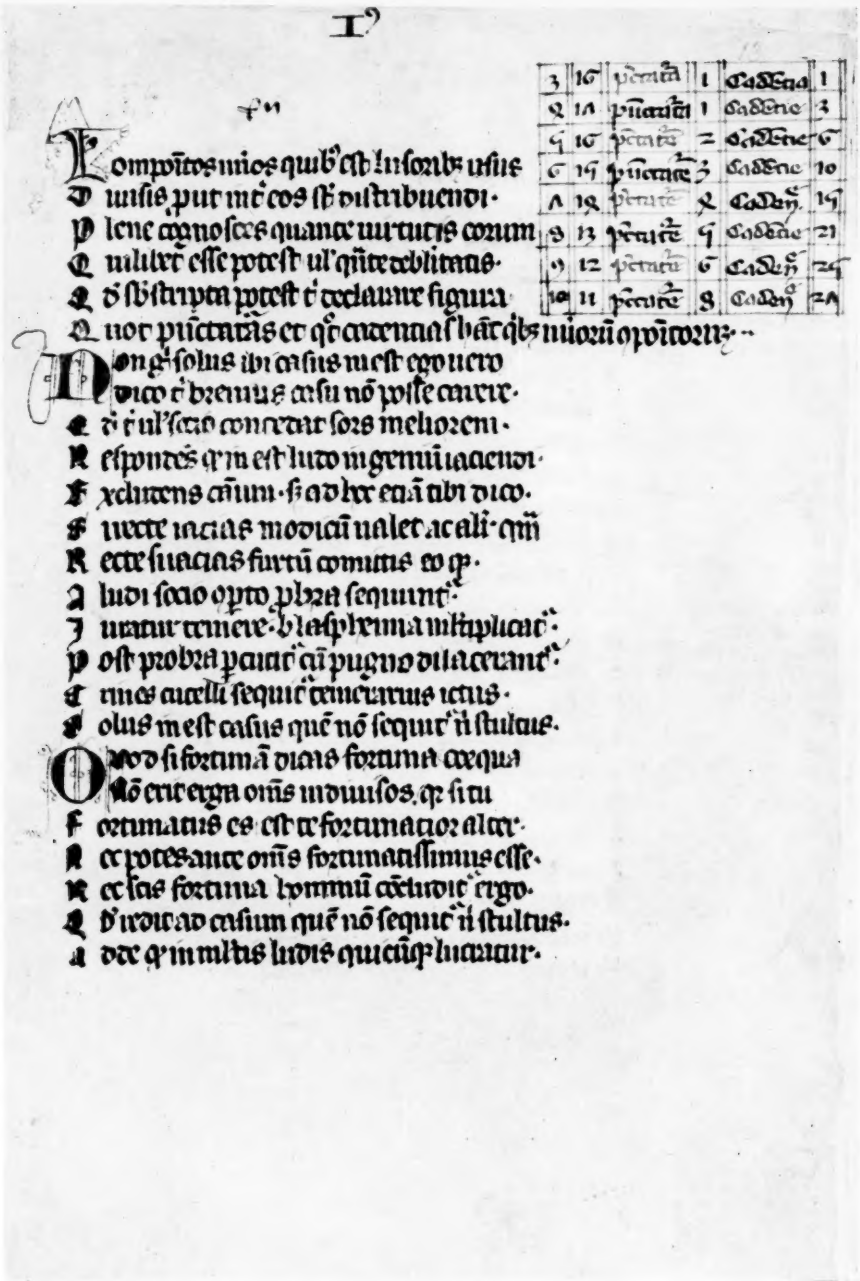
Qvot Punctaturas, et qvot Cadentias habeat qvilibet numerorū compoſitorum.

3	18	Punctatura	1	Cadentia	1
4	17	Punctatura	1	Cadentiæ	3
5	16	Punctaturæ	2	Cadentiæ	6
6	15	Punctaturæ	3	Cadentiæ	10
7	14	Punctaturæ	4	Cadentiæ	15
8	13	Punctaturæ	5	Cadentiæ	21
9	12	Punctaturæ	6	Cadentiæ	25
10	11	Punctaturæ	6	Cadentiæ	27

Fig. 1

18. From these passages it seems clear that by the end of the fifteenth century the foundations of a doctrine of chance was being laid. The necessary conceptualization of the perfect die and the equal frequency of occurrence of each face are explicit. The idea of attaching binomial coefficients to the possibilities with two or more dice in order to calculate their relative frequency of occurrence had occurred in the *De Vetula* but seems to have been lost to sight. Not until 1556 did Tartaglia publish the scheme now known (very unjustly) as Pascal's arithmetic triangle, and then not in a probabilistic context. Nevertheless, if it is the first step which counts, that step had already been made by A.D. 1500.





19. Although the pioneer step in the *De Vetula* seems to have been overlooked by later writers in the next two centuries, the idea of enumerating the ways of obtaining given scores when permutations were taken into account must have been rediscovered by the beginning of the sixteenth century; for Cardano's *De Ludo Aleae* contains the essential ideas and is dated on internal evidence as written about 1526.* Oddly enough, however, we find the first problems in probability (so far noticed in the records) in quite a different context.

20. Fra Luca dal Borgo, or Paccioli, was an itinerant teacher of mathematics whose *Summa de Arithmetica, Geometria, Proportioni et Proportionalità*, published in 1494, was widely studied in Italy. He considers a simple version of what later became known as the problem of points: A and B, playing at a fair game (not dice, but *balla*, presumably a ball game) agree to continue until one has won six rounds; but the match has to stop when A has won five and B has won three. How should the stakes be divided?

21. Paccioli makes very heavy weather of this, but his solution amounts to saying that the stakes should be divided in the proportion 5:3. The error was noted by Tartaglia in his monumental *General Trattato* of 1556 (which date, we may remark, is thirty years after Cardano says that he was in possession of the basic principles embodied in the *De Ludo Aleae*). Tartaglia was always glad to point out errors in Paccioli with an acid superiority which foreshadows many of the modern writings on probability and statistics. He would have been more justified on this occasion if the alternative solution which he propounds had been correct, which it is not. He points out that according to Paccioli's rule, if A had won one game and B none, A would take all the stakes, which is obviously unjust. He then argues that the difference between A's score (five) and B's score (three) being two, and this being one-third of the number of games needed to win (six), A should take one third of B's share and the total stake should be divided in the ratio 2:1. Or so I interpret his rather prolix discussion. It would appear that if A has x and B y games in hand when the total number required to win is z, Tartaglia's rule requires that A takes a proportion $\frac{1}{2}+(x-y)/(2z)$ of the stake.

22. Two years after the *Trattato* there appeared a short work by G. F. Peverone, *Due Brevi e Facili Trattati, il Primo d'Arithmetica, l'Altro di Geometria*. In the first of these Peverone considers a similar problem, without reference to other writers. A has won 7 and B 9 games in a match going to 10 games. He gives two examples, which are effectively the same, and argues in this way:

> A should put 2 crowns and B 12 crowns [or, equivalently, the stake should be divided in the proportion 1:6]. For if A, like B, had one game to go each would put two crowns [or divide the stakes in equal proportions]. If A had two games to go against B's one, he should put 6 crowns against B's two, because, by winning two games he would have won four crowns,† but with the risk of losing the second after winning the first; and with three games to go he should put 12 crowns because the difficulty and risk are doubled.

* Dr David informs me that she is convinced that Cardano obtained the substance of his work on gambling from other sources. This would be in accordance with Cardano's character, for he was not an originator in spite of his extensive knowledge and peculiar gifts. On the other hand, it must remain a conjecture until those sources can be traced.

† 'Se giuocassero a 1 giuoco, bastarebbero scutti 2; et a due giuochi 6, per che vincendo solo 2 giuochi guadagnarebbe scutti 4; ma questo sta con pericolo di perdere il secondo, vinto il primo: però deve guadagnare scutti 6, et a 3 giuochi scutti 12, per che si indoppia la difficoltà e pericolo.'

23. I think this must be one of the nearest misses in mathematics. As far as the second game the argument is correct. If B has one game to go and is staking two crowns, then for A:

with one game to go he stakes 2 crowns,
with two games to go he stakes $2+4=6$ crowns,
with three games to go he stakes $2+4+8=14$ crowns,

and so on. Peverone was perfectly well acquainted with geometrical progressions and uses the word *progressione* in one exposition of his answer to this problem. Having got as far as the staking of 6 crowns by A with two games to go, if he had only stuck to his own rule and considered the conditional probabilities of gain more closely, he would have solved this simple case of the problem of points, in essence, nearly a century before Fermat and Pascal.

24. Students of the modern versions of hazard and primero have been struck by the accurate judgement of probabilities which they embody. The chance of success of the first player at craps, for example, should be 1/2 and is actually 244/493; the relative values of flush and straight at poker are correct although intuitively it is not clear what the order should be. It seems, however, that this situation has been reached empirically and not by calculation. Cardano fortunately gives us an account of primero as known to him.* From the pack of 52 the eights, nines and tens were removed, leaving 40. Four were dealt to each player two at a time. The cards had individual values, two counting 12, three 13, four 14 and five 15; six counted 18 and seven 21; an ace 16 and court cards 10. There were five combinations:

(*a*) Numerus (two or three cards of the same suit);
(*b*) Primero (all cards of different suits);
(*c*) Supremus (the three cards 7, 6, ace in the same suit);
(*d*) Fluxus (four cards of the same suit);
(*e*) Chorus (all cards of the same denomination).

These were valued in that order, a primero beating a numerus and so forth. The categories do not overlap and if two players held the same combination the elder hand won, irrespective of suit.

Now the chances of these events, or rather the number of ways in which they can occur on random drawing, are

Chorus		10
Fluxus		840
Supremus		120
Primero		8,990
Two of a suit	54,000	
Three of a suit	14,280	
Two pairs	12,150	80,430
		90,390

* In the Middle Ages many games were played in several forms. For example, there were about a dozen different versions of chess. Primero as described by Cardano does not seem to have incorporated a draw for new cards. Sir John Harington, in the reign of Elizabeth I, refers to a later version called 'prime' which does. Robert Greene, in 1591, makes a character say 'what will you play at, primero, primo visto, sant, one-and-thirty, new cut, or what shall be the game?' Shakespeare also mentions primero. I have seen it stated that the game (more properly primera) was imported from Spain on the occasion of the marriage of Mary Tudor with Philip II. The terms used in it certainly suggest a Spanish origin, but I do not see why the marriage of Henry VIII with Catharine of Aragon could not have been the occasion, or, indeed, that a specific occasion need be invoked.

In other words, the relative value of the fluxus and the supremus in Cardano's version were in the inverse order of their probabilities. At what point the present (correct) order in the modern game of poker emerged, I do not know, but it seems to have been before anyone was in a position to calculate the chances and persuade his fellows on the basis of mathematics that the orders should be reversed. In my opinion relative chances were all reached on the basis of intuition or trial and error in the games played up to the middle of the seventeenth century.

25. It seems clear that in fifteenth-century Italy the basic problems of chance in gaming had been raised and some small progress made towards solving them. A more thorough examination of the Italian mathematical books of the period may reveal further evidence on the point. One suspects that some of the simpler problems were circulated as a kind of puzzle, just as they are at the present day, without becoming of any recognized scientific importance. Galileo in his fragment *Sulla Scoperta dei Dadi*, written some time before 1642 (the date of his death), gives a complete solution of a problem in direct probability by correct enumeration of all possibilities, and he writes as if the problem were a new one, mentioning no previous authors.* Nevertheless, if Cardano's treatise is to be correctly assigned to 1526 the ideas must have been current for a century before Galileo wrote. It would appear that a calculus of probability not only was late in developing but that, once begun, it progressed exceedingly slowly.

26. Before we consider the reasons for this, something remains to be said about developments in France in the first half of the seventeenth century. The cradle of the probability calculus was undoubtedly, in my opinion, in Italy. From the fourteenth century, however, there were close connexions between France and Italy of a political as well as a geographical kind and an intellectual movement in one often generated a sympathetic movement in the other. The invasion of Italy by Charles VIII in 1494, though militarily and politically a failure, is generally regarded as a useful piece of intellectual cross-fertilization. Undoubtedly, a great many Italian works of art and ideas found their way to France with the remnant of Charles's army, although I doubt whether a copy of Paccioli's book was amongst them. In this case also a search among French books on mathematics written between A.D. 1400 and 1650 might prove to be very instructive.

27. The lack of written references to problems in the probability calculus is not necessarily indicative of a lack of contemporary interest. Knowledge of chances was so rudimentary that any capacity to gauge them accurately in play was worth a good deal of money. Huyghens, visiting France in 1657, found intense interest being taken in the doctrine of chances among mathematicians but encountered also a certain coyness about the disclosure of results. This was presumably due to fear of anticipatory publication rather than loss of income. Huyghens, being the man he was, merely worked out the theory for himself. A Latin translation of his little book, *De Ratiociniis in Ludo Aleae*, printed by van Schooten in 1657, was the first book published on the probability calculus and exercised a profound influence on James Bernoulli and Demoivre.

28. Now we come to the most interesting question of this period. Why was it that the calculus of probabilities was so long in emerging? We cannot suppose that the Greeks were incapable of making the necessary generalizations, even if they were hampered in working

* This is not a very weighty consideration. Early writers on probability, like those of the present day, often failed to mention their indebtedness to their predecessors. Laplace was notoriously bad at it.

out the details by their arithmetic and algebra. The same is true of the Arabs and of the early medieval Europeans. Dr David has suggested that imperfections in the dice may have something to do with it, but I cannot believe that this was a major reason. Some of the dice were, in fact, quite well made. The races which built the Parthenon, Trajan's Column, St Sophia and Notre Dame were quite capable of turning out a few cubes as good as any of those in current use. Nor do I think backward mathematical notation had much to do with it. We have seen that the partitional falls of dice were counted without difficulty in the tenth century. Four other possibilities are worth examining:

(*a*) the absence of a combinatorial algebra (or at any rate, of combinatorial ideas);
(*b*) the superstition of gamblers;
(*c*) the absence of a notion of chance events;
(*d*) moral or religious barriers to the development of the idea of randomness and chance.

29. Combinatorial algebra does not seem to have been cultivated by the ancients. Interest in it awoke in the sixteenth and seventeenth centuries. Leibniz published a tract *De Arte Combinatoria* in 1660 and Wallis a *De Combinationibus Alternationibus et Partibus Aliquotis Tractatus* in 1685. Doubtless the essential ideas could be traced back a good deal earlier.* Thus, when the calculus of probability was really under way, a combinatorial algebra lay ready to hand. Nevertheless, it seems to me, the absence of such an algebra cannot be held to account for the late emergence of the doctrine of chance. Cardano managed well enough without it and Galileo's enumeration of the 216 ways of throwing three dice is perfect, though apparently based only on arithmetical methods.

30. The superstition of gamblers is well known and has been remarked upon by many early writers. If men were logical and observant beings, one would have deemed it impossible for any person to engage very much at play and, at the same time, to believe that the favours of fortune were distributed unequally in the long run. But it seems quite possible for a player to believe two incompatible propositions; and with sufficient ingenuity, I suppose, it is also possible to reconcile a belief in the law of large numbers with a belief that the luck will change if one takes a different chair. One can say much on this subject. I am content to record the opinion that, although the psychology of the gambler may have done something to hinder the development of a concept of probabilistic law, it cannot have prevented the leading minds of the age from arriving at such a concept.

31. If we discount such factors as ill-made dice, indifferent mathematical expertise, superstition and so forth; and if we agree that play with dice and cards were so prevalent as to arouse general interest among intelligent people; then we seem driven to the conclusion that the late emergence of the probability calculus was due to some more fundamental factor. The very notion of chance itself, the idea of natural law, the possibility that a proposition may be true and false in fixed relative proportions, all such concepts are nowadays so much part of our common routines of thought that perhaps we forget that they were not so to our ancestors. It is in basic attitudes towards the phenomenal world, in religious and moral teachings and barriers, that I incline to seek for an explanation of the delay. Mathematics never leads thought, but only expresses it.

* The origins of combinatorial algebra would themselves make an interesting historical study. Wallis, at the age of 25, established a reputation for himself by deciphering Royalist letters intercepted during the Civil War. Bacon, in the reign of Elizabeth I, also took a keen interest in cryptography. Both men used ciphers based on combinations of symbols.

32. The Greeks and the Romans (so far as one can make summary statements about races whose members held such differing views) seem, on the whole, to have regarded the world as partly determined by chance. Gods and goddesses had influence over the course of events and, in particular, could interfere with the throwing of dice; but they were only higher beings with superhuman powers, not omnipotent entities who controlled everything. And the vaguer deities, Fortuna, the Fates and Fate itself appear to modern eyes more in the retributive role of a personified guilty conscience than as masters of the universe. The situation was radically changed by Christianity. For the early fathers of the Church the finger of God was everywhere. Some causes were overt and some were hidden, but nothing happened without cause. In that sense nothing was random and there was no chance. 'Nos eas causas', says St Augustine, 'quae dicuntur fortuitae (unde etiam fortuna nomen accepit) non dicimus nullas, sed latentes; easque tribuimus vel veri Dei, vel quorumlibet spirituum voluntati.' This view prevailed also in medieval times. Thomas Aquinas, arguing that everything is subject to the providence of God, mentions explicitly the objection that, if such were the case, hazard and luck would disappear. He replies that there are universal and particular causes; a thing can escape the order of a particular cause but not of a universal cause; and so far as it escapes it is said to be fortuitous with respect to that cause. St Thomas has an Aristotelean view of primary and secondary causes but we need not follow closely his struggles with the problems of causality, predestination and free-will. He reflected the spirit of his age, wherein God and an elaborate hierarchy of His ministers controlled and fore-ordained the minutest happening; if anything seemed to be due to chance that was our ignorance, not the nature of things.*

33. St Thomas is sometimes quoted as having expressed himself in favour of a frequency theory of probability, but, in my opinion, this rests on a source of confusion which it may be useful to remove in passing. Throughout this article I have been speaking of the doctrine of chances (which Demoivre translated as *Mensura Sortis*), not probability in the wider sense. Early writers used *probabilitas* with a different meaning, as relating to the degree of doubt with which a proposition is entertained. At the outset of our science the two things were distinct and it is a pity that they have not remained so and that our language has tended to confuse them. It seems to have been James Bernoulli who first thought of applying the doctrine of chances to the art of conjecture; and although we find applications to the assessment of the credibility of witnesses as early as 1697, it was not until Bayes' time (1763) that it was also applied to the acceptability of hypotheses. The resulting confusion, as is well known, has existed ever since and at the present time seems, if anything, to be getting worse. If any justification for the study of the history of probability and statistics were required, it would be found simply and abundantly in this, that a knowledge of the development of the subject would have rendered superfluous much of what has been written about it in the last thirty years.

34. Aquinas does not give a definition of *probabilitas*, but refers to Aristotle; 'Probabilia sunt quae videntur omnibus, aut plerisque, aut sapientibus, et his vel omnibus vel

* The same idea, of course, has come down to modern times in a line of direct descent but in a less deistic form; e.g. Spinoza, writing in 1677, says 'for a thing cannot be called contingent unless with reference to a deficiency in our knowledge'; D'Alembert, in 1750, says: 'il n'y a point de hasard à proprement parler mais il y a son équivalent: l'ignorance où nous sommes des vraies causes des événements.' More recently Paul Lévy in 1939: 'Nous pensons, quoique depuis les travaux d'Heisenberg d'éminents savants ne soient pas de cet avis, que la notion du hasard est une notion que le savant introduit parce qu'elle est commode et féconde, mais que la nature ignore.'

plurimis maxime nobilibus et probatis.' St Thomas himself regarded *probabilitas* as a quality which gave rise to an opinion. He says explicitly that it admits of degrees. Of chance (casus), he says: 'Ea quae accidunt semper vel frequenter non sunt casualia neque fortuita, sed quae accidunt in paucioribus.' And again 'Sicut in rebus naturalibus in his quae ut in pluribus agunt, gradus quidam attenditur quia quanto virtus naturae est fortior, tanto rarius deficit a suo effectu, ita et in processu rationis qui non est cum omnimoda certitudine, gradus aliquis invenitur, secundum quod magis et minus ad perfectum certitudinem acceditur.' As I understand the position, St Thomas recognized that *probabilitas* preceded certainty in the formation of knowledge; and that the frequency of events had something to do with the 'fortuitous' nature of the causality and the relative intensity of the underlying cause. But I cannot see in his writings an explicit statement that *frequentia* increased *probabilitas* or that the two were very closely related. The point really needs a deeper study than I felt worth while to give it, but it seems plain to me that the doctrine of chances was not present to his mind when *probabilitas* was under discussion.

35. A good case can be made for the thesis that the religious attitude of the times discouraged by implication the development of a study of random behaviour. Even this does not entirely satisfy me as the complete explanation, but I think it very likely that before the Reformation the feeling that every event, however trivial, happened under Divine providence may have been a severe obstacle to the development of a calculus of chances. It seems to have taken humanity several hundred years to accustom itself to a world wherein some events were without cause; or, at least, wherein large fields of events were determined by a causality so remote that they could be accurately represented by a non-causal model.* And, indeed, humanity as a whole has not accustomed itself to the idea yet. Man in his childhood is still afraid of the dark, and few prospects are darker than the future of a universe subject only to mechanistic law and to blind chance.

Whatever the reasons may be, it appears undeniable that the doctrine of chances took a remarkably long time to develop. Once launched, of course, it proceeded very rapidly; there is only a hundred years between Bernoulli's *Ars Coniectandi* and Laplace's *Traité*. But the results of that century of discovery required several thousand years of germination. Until more intensive research may have been able to lay bare the early essays and modes of thought of scientists in the fifteenth, sixteenth and seventeenth centuries the birth process of the probability calculus must remain somewhat enigmatic.

36. I have to express acknowledgements to several colleagues for help in tracing references; especially to Prof. Corrado Gini, to Father Dionisio Pacetti, O.F.M., whose authoritative knowledge of the works of St Thomas Aquinas was put freely at my disposal, to Prof. Sixtos Rios for some information about the game of primero, to Prof. W. Rose and to Mr G. Woledge for some references to early German and French gaming, and to

* As is well known, the assurance of lives was forbidden in several countries under Roman Catholic influence in the sixteenth century; e.g. in the statutes of Genoa for 1588: 'sine licentia Senatus non possunt fieri securitate...super vitam Pontificis neque super vitam...aliorum dominorum aut personarum ecclesiasticarum.' It is rather remarkable, but I suppose accidental, that nearly all the chief early writers on the probability calculus were subject to Catholic persecution. Cardano and Galileo were both victims of the Inquisition. The Bernoulli family were exiled in Switzerland, having been driven from Antwerp by Spanish persecution in the Netherlands. De Moivre lived in England because of the revocation of the Edict of Nantes. Pascal was a member of Port-Royal, though he did not live to see Jansenism driven into exile. Fermat and Montmort escaped; but Fermat lived in the provinces and published nothing on probability, and Montmort also lived a retired life.

Dr F. N. David, with whom I have had many discussions on this fascinating subject and who read this article in manuscript.

REFERENCES

DAVID, F. N. (1955). Dicing and gaming (a note on the history of probability). *Biometrika*, **42**, 1.
LIBRI, C. D. S. (1838–41). *Histoire des Sciences Mathématiques en Italie*, 2 vols. Paris.

APPENDIX

Extract from 'De Vetula'

Forte tamen dices, quosdam praestare quibusdam
Ex numeris, quibus est lusoribus usus, eo quod
Cum decius sit sex laterum, sex & numerorum
Simplicium, tribus in deciis sunt octo decemque,
Quorum non nisi tres possunt deciis superesse.
Hi diversimode variantur, & inde bis octo
Compositi numeri nascuntur, non tamen aequae
Virtutis, quoniam majores atque minores
Ipsorum raro veniunt, mediique frequenter,
Et reliqui, quanto mediis quamvis propiores,
Tanto praestantes, & saepius advenientes.
His punctatura tantum venientibus una,
Illis sex, aliis mediocriter inter utrosque,
Sicut sint duo majores, totidemque minores,
Una quibus sit punctatura, duoque sequentes,
Hic major, minor ille, quibus sit bina duobus.
Rursum post istos sit terna, deinde quaterna,
Quinaque, sicut eis succedunt appropiando
Quattuor ad medios, quibus est punctatio sena,
Quae reddet leviora tibi subjecta tabella.

Hi sunt sex & quinquaginta modi veniendi,
Nec numerus minor esse potest, vel major, eorum.
Nam quando similes fuerint sibi tres numeri, qui
Jactum componunt quia sex componibiles sunt,
Et punctaturae sunt sex, pro quolibet una.
Sed cum dissimilis aliis est unus eorum,
Atque duo similes, triginta potest variari
Punctatura modis, quia, si duplicaveris ex se
Quemlibet, adjuncto reliquorum quolibet, inde
Producens triginta, quasi sex quintuplicabis.
Quod si dissimiles fuerint omnino sibi tres,
Tunc punctaturas viginti connumerabis.
Hoc ideo, quia continui possunt numeri tres
Quattuor esse modis; discontinui totidem: sed
Si duo continui fuerint, discontinuusque
Tertius invenies hinc tres bis, & inde duos ter:
Quod tibi declarat oculis subjecta figura.

Rursum sunt quaedam subtilius inspicienti
De punctaturis, quibus una cadentia tantum est;
Suntque; quibus sunt tres aut sex quia schema cadendi
Tunc differe nequit, quando similes fuerint tres
Praedicti numeri. Si vero sit unus eorum
Dissimilis, similisque duo, tria schemata surgunt,
Dissimili cuicunque superposito deciorum.
Sed si dissimiles sunt omnes, invenies sex
Verti posse modis, quia, quemlibet ex tribus uni

Cum dederis, reliqui duo permutant loca; sicut
Punctaturarum docet alternatio. Sicque
Quinquaginta modis et sex diversificantur
In punctaturis, punctaturaeque ducentis
Atque bis octo cadendi schematibus, quibus inter
Compositos numeros, quibus est lusoribus usus,
Divisis, prout inter eos sunt distribuendi
Plene cognosces, quantae virtutis eorum
Quilibet esse potest, seu quantae debilitatis:
Quod subscripta potest tibi declarare figura.

ON ESTIMATING THE LATENT AND INFECTIOUS PERIODS OF MEASLES

I. FAMILIES WITH TWO SUSCEPTIBLES ONLY

By NORMAN T. J. BAILEY

Design and Analysis of Scientific Experiment, 6 Keble Road, Oxford

1. Introduction

The analysis of the household distribution of cases of measles by means of chain-binomial models has been comparatively successful (see Bailey (1955) for discussion and bibliography). At the same time considerable variation in the time interval between successive cases has usually been observed, and this sometimes leads to difficulties in identifying the links of the chain. An attempt was therefore made to produce a model which would take these variations into account. The simplest feasible arrangement is to assume that after the receipt of infection there follows a latent period which is approximately normally distributed. Then comes a period of infectiousness which is effectively terminated after a constant time by the appearance of symptoms and removal of the individual concerned from circulation. The latent and infectious periods taken together constitute what is usually called the incubation period. Arguments in favour of this model have been set out in detail elsewhere (see Bailey (1954, 1955), to which reference should be made for a fuller discussion). Simple estimates of the four parameters involved were given for families with two susceptibles only. However, at least two of the large sample variances were high, and there was considerable doubt as to the efficiency achieved. The purpose of the present paper is accordingly to develop a maximum-likelihood scoring procedure which will give efficient estimates and also make available a goodness-of-fit test.

2. Mathematical model and sampling distributions

Let us suppose that the latent period, x, is normally distributed with mean m and variance σ^2, while the ensuing infectious period is of constant length a. Infection of the second susceptible during this time is taken to be a Poisson process such that the chance of contracting the disease in time dt is $\lambda\,dt$.

Suitable material for analysis (kindly made available to me by Dr R. E. Hope Simpson) is shown in Table 1. This is based on families with two susceptible children under 15 years of age and at least one case of measles, and was taken from the Cirencester area over the years 1946–52. The distribution appears to involve two distinct parts which overlap a little. Distribution A containing A families is considered to arise from both susceptibles having been simultaneously infected by an outside contact. The B families in the B-distribution are taken to be examples of cross-infection within the family. We shall assume for the time being that observations can be allotted to the correct distribution with complete accuracy. In practice an arbitrary decision (as in Table 1) may have to be made about borderline cases. A refinement of analysis is to introduce and estimate an additional parameter, the prior chance that a family belongs to, say, distribution A. The kind of procedure required is outlined below in §6. There are C families with only one case, and a total of $N = A+B+C$.

Let w be the variable for distribution A. Since this is the absolute difference between two independent latent periods the frequency distribution of w is

$$f(w) = \frac{1}{\sigma\sqrt{\pi}} e^{-w^2/(4\sigma^2)} \quad (0 \leqslant w < \infty). \tag{1}$$

Next, consider the $B+C$ families with either both a primary and a secondary case (B) or a single primary case only (C). The chance of the second susceptible escaping infection by the first case during the infectious period a is $e^{-\lambda a}$. Hence the probability of the observed numbers, B and C, given $B+C$, is

$$\binom{B+C}{B} e^{-C\lambda a}(1-e^{-\lambda a})^B. \tag{2}$$

Further, the distribution of the epoch τ, measuring from the beginning of the infectious period, at which infection of the second case actually takes place, is evidently

$$f(\tau) = \lambda e^{-\lambda\tau}(1-e^{-\lambda a})^{-1} \quad (0 \leqslant \tau \leqslant a). \tag{3}$$

Let distribution B arise from a variable $z = x+\tau$, where of course x has the frequency function

$$f(x) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\{-(x-m)^2/(2\sigma^2)\} \quad (-\infty < x < \infty). \tag{4}$$

The frequency distribution of z is obtained by using (3) and (4) to write down the joint frequency distribution for x and τ, replacing x by $z-\tau$, and integrating out τ. This gives

$$f(z) = \frac{\lambda e^{-\lambda(z-m-\frac{1}{2}\lambda\sigma^2)}}{1-e^{-\lambda a}} \int_{u'}^{u} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}t^2}\, dt, \tag{5}$$

where

$$u = \sigma^{-1}\{z-(m+\lambda\sigma^2)\} \quad \text{and} \quad u' = u - a\sigma^{-1}. \tag{6}$$

We shall also need the sample mean, $\bar{z}$, and variance, v, of the B-distribution, together with the observed second moment about the origin, $V = \Sigma w^2/A$, for the A-distribution.

3. Maximum-likelihood scoring

Using the three frequency functions given in (1), (2) and (5) above we can proceed in the usual way to derive maximum-likelihood scores and information functions for the parameters λ, a, m and σ. In order to do this as concisely as possible let us first write

$$R(z) \equiv \int_{u'}^{u} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}t^2}\, dt, \tag{7}$$

and

$$T_\theta = \frac{1}{R}\frac{\partial R}{\partial \theta} \quad (\theta = \lambda, a, m, \sigma). \tag{8}$$

Amalgamating the three contributions to any score then gives quite simply

$$\left.\begin{aligned} S_\lambda \equiv \partial L/\partial\lambda &= B(m-\bar{z}+\lambda^{-1}+\lambda\sigma^2) - Ca + \sum_z T_\lambda, \\ S_a \equiv \partial L/\partial a &= -C\lambda + \sum_z T_a, \\ S_m \equiv \partial L/\partial m &= B\lambda + \sum_z T_m, \\ S_\sigma \equiv \partial L/\partial\sigma &= A\sigma^{-1}(\tfrac{1}{2}V\sigma^{-2}-1) + B\lambda\sigma^2 + \sum_z T_\sigma, \end{aligned}\right\} \tag{9}$$

where L is as usual the log likelihood. The quantities T_θ are most easily calculated from the functions

$$\left.\begin{aligned} Q(x) &= \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2} \quad \text{and} \quad P(x) = \int_{-x}^{x} Q(t)\,dt \quad (x \geqslant 0) \\ &= -P(|\,x\,|) \quad (x \leqslant 0), \end{aligned}\right\} \tag{10}$$

using the tables published by the New York W.P.A. (1942). We then have

$$\left.\begin{aligned} &R = \tfrac{1}{2}(P-P'), \quad \partial R/\partial\lambda = -\sigma(Q-Q'), \\ &\partial R/\partial a = \sigma^{-1}Q', \quad \partial R/\partial m = -\sigma^{-1}(Q-Q'), \\ &\partial R/\partial\sigma = -\sigma^{-1}(uQ-u'Q') - 2\lambda(Q-Q'), \\ &\text{where} \qquad P \equiv P(u), \quad P' \equiv P(u'), \quad Q \equiv Q(u), \quad Q' \equiv Q(u'). \end{aligned}\right\} \tag{11}$$

The T_θ are then obtained from (8), (10) and (11).

The derivation of information functions also goes in a straightforward way. A minor point worth mentioning is that if we differentiate T_θ with respect to one of the parameters, say ϕ, we obtain

$$-T_\theta T_\phi + \frac{1}{R}\frac{\partial^2 R}{\partial\phi\,\partial\theta}.$$

The expectation of $R^{-1}\partial^2 R/\partial\phi\,\partial\theta$ is easily found, since when multiplying by the frequency function in (5) the factors R in numerator and denominator cancel, and the integration with respect to z gives no special difficulty. With $T_\theta T_\phi$, on the other hand, the integrand involves a factor R^{-1}, and it seems best to leave these terms as observed quantities. In any case the individual values of T_θ and T_ϕ for each z have already been calculated in finding the scores. The information matrix, I, for the parameters λ, a, m and σ, in that order, then turns out to be

$$\begin{bmatrix} \Sigma T_\lambda^2 - \bar{B}(\lambda^2\sigma^4 + \sigma^2 - \lambda^{-2}), & \Sigma T_\lambda T_a + \dfrac{\bar{B}(1+\lambda^2\sigma^2)}{e^{\lambda a}-1}, & \Sigma T_\lambda T_m - \bar{B}(1+\lambda^2\sigma^2), & \Sigma T_\lambda T_\sigma - \bar{B}\lambda^3\sigma^3 \\ & \Sigma T_a^2 + \dfrac{\bar{B}\lambda^2}{e^{\lambda a}-1}, & \Sigma T_a T_m + \dfrac{\bar{B}\lambda^2}{e^{\lambda a}-1}, & \Sigma T_a T_\sigma + \dfrac{\bar{B}\lambda^3\sigma}{e^{\lambda a}-1} \\ & & \Sigma T_m^2 - \bar{B}\lambda^2, & \Sigma T_m T_\sigma - \bar{B}\lambda^3\sigma \\ & & & \Sigma T_\sigma^2 - \bar{B}\lambda^4\sigma^2 + 2A\sigma^{-2} \end{bmatrix}, \tag{12}$$

where

$$\bar{B} = (B+C)(1-e^{-\lambda a}). \tag{13}$$

Writing $\mathbf{S}$ for the vector of scores calculated at trial values given by the vector $\boldsymbol{\theta}$, we calculate approximate maximum-likelihood values, $\boldsymbol{\theta}_1$, given by

$$\boldsymbol{\theta}_1 = \boldsymbol{\theta} + \mathbf{I}^{-1}\mathbf{S}, \tag{14}$$

as is well known. The procedure is then repeated using $\boldsymbol{\theta}_1$ for the trial values until sufficient accuracy is obtained.

To obtain initial trial values it is convenient to use the approximate estimates given in a previous paper (Bailey, 1955), namely,

$$\left.\begin{aligned} \lambda &= (1-T)^{\frac{1}{2}}U^{-\frac{1}{2}}, \\ a &= \lambda^{-1}\log(1+Y), \\ m &= \bar{z} - \lambda^{-1} + aY^{-1}, \\ \sigma &= (\tfrac{1}{2}V)^{\frac{1}{2}}, \end{aligned}\right\} \tag{15}$$

where

$$\left.\begin{aligned} T &= Y^{-1}(1+Y^{-1})\{\log(1+Y)\}^2, \\ U &= v-\tfrac{1}{2}V, \\ Y &= B/C. \end{aligned}\right\} \tag{16}$$

If it should happen that in any set of data A were small or zero then no satisfactory initial estimate of σ would be available from (15). However, it is still possible to obtain rough estimates by setting B equal to its expectation, leading to

$$\lambda a = \log(1+Y) = F, \tag{17}$$

and putting the first three sample cumulants in the B-distribution equal to their expectations, viz.

$$\left.\begin{aligned} \kappa_1 &\equiv m+\lambda^{-1}-a(e^{\lambda a}-1)^{-1} &&= \bar{z}, \\ \kappa_2 &\equiv \sigma^2+\lambda^{-2}-a^2e^{\lambda a}(e^{\lambda a}-1)^{-2} &&= v, \\ \kappa_3 &\equiv 2\lambda^{-3}-a^3e^{\lambda a}(e^{\lambda a}+1)(e^{\lambda a}-1)^{-3} &&= m_3. \end{aligned}\right\} \tag{18}$$

Using (17) and (18), we can solve to the required estimates successively as follows:

$$\left.\begin{aligned} a^3 &= m_3[2F^{-3}-Y^{-3}(1+Y)(2+Y)]^{-1}, \\ \sigma^2 &= v-a^2[F^{-2}-Y^{-2}(1+Y)], \\ m &= \bar{z}-a(F^{-1}-Y^{-1}), \\ \lambda &= Fa^{-1}. \end{aligned}\right\} \tag{19}$$

4. Illustrative example

Let us now apply the foregoing maximum-likelihood scoring procedure to the data exhibited in Table 1. Examination of the observed frequencies shows immediately that in attempting to fit any reasonably smooth curve to B we are likely to be troubled by an apparent excess of observations for days 7 and 14 (with corresponding deficits for adjacent days). This could be due to a small unconscious bias towards an integral number of weeks. A similar situation is not uncommon in other fields, e.g. interviewees may show a preference for ages ending in 0 or 5, and measurements of blood pressure sometimes show a marked excess of readings which are even multiples of 10. One way of offsetting such bias is to pool the frequencies for, say, days 6, 7 and 8 in one group, and for 13, 14 and 15 in another group, when carrying out a goodness-of-fit test.

Preliminary estimates were obtained from (15). These have already been given in an earlier paper (Bailey, 1955) together with their large sample variances, and are

$$\left.\begin{aligned} \lambda &= 0{\cdot}203 \pm 0{\cdot}083, \\ a &= 8{\cdot}13 \pm 3{\cdot}28 \text{ days}, \\ m &= 7{\cdot}94 \pm 0{\cdot}87 \text{ days}, \\ \sigma &= 1{\cdot}32 \pm 0{\cdot}17 \text{ days}. \end{aligned}\right\} \tag{20}$$

The estimates of both λ and a appear to have rather low precision, and we should not be surprised if the maximum-likelihood values are appreciably different and much more accurate. After one cycle of successive approximation the second set of estimates did in fact show large changes. When the third set of values was obtained the information matrix was recalculated, and convergence was then rapid. Stability was practically achieved with the

fourth approximation—a final fifth stage gave only very small further corrections. The maximum-likelihood estimates found in this way are

$$\left.\begin{aligned} \hat{\lambda} &= 0{\cdot}256 \pm 0{\cdot}032, \\ \hat{a} &= 6{\cdot}57 \pm 0{\cdot}76 \text{ days}, \\ \hat{m} &= 8{\cdot}58 \pm 0{\cdot}32 \text{ days}, \\ \hat{\sigma} &= 1{\cdot}77 \pm 0{\cdot}13 \text{ days}. \end{aligned}\right\} \tag{21}$$

There is thus a striking gain in efficiency compared with the preliminary estimates appearing in (20), and the additional labour in computation is evidently worth while.

When carrying out the usual goodness-of-fit test it is convenient to use the last approximation but one, if sufficiently accurate, since we need not then recalculate the $R(z)$ in computing the function given by (5). In practice, of course, the data will normally be grouped in units of 1 day, as in Table 1. The fitted values for distribution B were obtained

Table 1. *Observed and expected values for Hope Simpson's data on measles in families with two susceptibles*

Time interval between 2 cases in days	Observed no. of families			Expected no.	
	A	*B*	Total		
0	5	.	5	4·67	
1	13	.	13	8·58	
2	5	.	5	6·73	
3	4	.	4	4·53	
4	2	1	3 } 5	2·78 } 5·03	
5	.	2	2 }	2·25 }	
6	.	4	4	3·97	
7	.	11	11	8·85	
8	.	5	5	16·63	
9	.	25	25	24·72	
10	.	37	37	29·44	
11	.	38	38	29·28	
12	.	26	26	25·44	
13	.	12	12	19·99	
14	.	15	15	14·28	
15	.	6	6	9·02	
16	.	3	3	4·82	
17	.	1	1	2·09	
18	.	3	3 } 8	0·71 } 7·84	
19	.	.	. }	0·18 }	
20	.	.	.	0·04	
21	.	1	1	0·00	
Sub-totals	29	190	219	219·00	
One case only (*C*)			45	44·11	
Primary and secondary (*B*)			190	190·89	
Overall total ($A+B+C$)			264	264	

merely by calculating $f(z)$ at the mid-point of each interval, as the small additional accuracy resulting from integration over the interval was thought not worth the extra labour of computation. On the other hand, integrated values for distribution A are immediately available from the tabulated values of $P(x)$ in (10). It can be seen from Table 1 that the agreement between observed and expected values is on the whole quite good. For the usual goodness-of-fit test the classes bracketed together have been pooled so as to avoid small expectations. There are sixteen classes for the combined A- and B-distribution, giving 15 d.f.; and two classes, giving 1 d.f., for the numbers B and C. From the total of 16 d.f. we must remove 4 to allow for the parameters estimated. We find an overall χ^2 of 20·3 on 12 d.f. As the 5 % point is at 21·0, we can regard the fit as just adequate. Actually, we have already remarked the possibility of unconscious bias in the records producing local peaks at 7 and 14 days, and suggested that it could be minimized by amalgamating the frequencies for 6, 7 and 8 days in one group, and 13, 14 and 15 in another. When this is done we obtain a χ^2 of 12·9 on 8 d.f., which is entirely satisfactory since the 10 % point is at 13·4.

Table 2. *Efficiencies (in percentage) of estimates with parts of data absent*

Data available	λ	a	m	σ
All three sources present	100	100	100	100
No double primaries ($A = 0$)	79	70	77	60
No single cases ($C = 0$)	23	100	83	99
B-distribution only ($A = 0 = C$)	22	70	68	59

It is also of some interest to see what would happen if certain portions of the data were missing. There may, for example, be no families with a double primary, i.e. $A = 0$. All we have to do is to remove the term $2A\sigma^{-2}$ from $I_{\sigma\sigma}$ in (12) before inverting the information matrix. Alternatively, there may be no record of C, the number of families with a single case only. No information about λ and a is then available from the frequency distribution in (2). Accordingly, we must remove from $I_{\lambda\lambda}$, $I_{\lambda a}$ and I_{aa} the contributions βa^2, $\beta a\lambda$ and $\beta\lambda^2$, respectively, where $\beta = (B+C)(e^{\lambda a}-1)^{-1}$. Again, both these items may be missing and we may only have data on families with a primary and a secondary case. These results are summarized in Table 2, which shows the appropriate efficiencies derived by comparing the variances of the estimates in each case with the 'best' values given by the squares of the standard errors appearing in (21). It is worth noticing that reasonably efficient estimates of all four parameters can be obtained in the absence of double primary data, but that knowledge of the number of families with a single case is essential for an efficient determination of λ. Without the B-distribution, of course, little information of value would be forthcoming.

5. Effect of variations in λ

It has been shown elsewhere (Bailey, 1953) that, so far as measles data for Providence, Rhode Island, were concerned, chain binomials gave a satisfactory goodness-of-fit when analysed stage by stage only on the assumption that the chance of cross-infection, p, varied between families. In the notation of this paper we have

$$p = 1 - e^{-\lambda a}. \tag{22}$$

It follows that we ought in the present context to consider the possibility of variations in λ. We could hardly expect to make any precise estimate of parameters in the distribution of λ from families with only two susceptibles, but it is worth making a rough estimate of the likely consequences. We can do this by calculating the expectation, with respect to appropriate variations in λ, of the expected frequencies in the model. Now the distribution chosen for p was

$$dF = \frac{1}{B(x,y)} p^{x-1}(1-p)^{y-1}\,dp \quad (0 \leqslant p \leqslant 1), \tag{23}$$

and the pooled estimates of x and y, based on households of three and four, were 1·18 and 0·28 respectively. The mean and variance of p were thus

$$\left.\begin{aligned} \bar{p} &= x/(x+y) = 0{\cdot}81, \\ v_p &= xy/\{(x+y)^2(x+y+1)\} = 0{\cdot}063. \end{aligned}\right\} \tag{24}$$

Using the estimates of λ and a in (21) we see that the value of p given by (22) is about 0·81, the same as the estimate in (24) taken from the Providence data. If we now suppose p to vary about this mean value the expected frequencies for the observations B and C will remain unchanged. The A-distribution will also be the same since it involves only σ. However, the B-distribution given in (5) will be modified, and may be replaced, approximately, by

$$f(z) + \frac{v_p}{2}\frac{\partial^2 f}{\partial p^2} = f(z) + \frac{v_p e^{2\lambda a}}{2a^2}\left(a\frac{\partial f}{\partial \lambda} + \frac{\partial^2 f}{\partial \lambda^2}\right), \tag{25}$$

where we have simply expanded in powers of $\delta p = p - \bar{p}$ and have taken expectations, neglecting terms of third and higher order in δp. It may be noted that the additional quantities on the right-hand side of (25) are all easily calculated from those already available. It turns out that the net result is to flatten and displace slightly towards the origin the peak of the fitted curve. The χ^2 values are a little higher. The data grouped so as to minimize the peaks at 7 and 14 days give a χ^2 of 14·3 on 8 d.f., which is still satisfactory. Without this adjustment we obtain 21·5 on 12 d.f., which is just significant at the 5 % level. However, the variations in p (or λ) are unlikely to be as large as this in Hope Simpson's data, which we should expect to be more homogeneous than the Providence material. We are entitled to conclude that our results would probably not be appreciably influenced by only moderate variations in the chance of infection.

A somewhat similar analysis can be undertaken in respect of variations in the length of the infectious period, a, but this is more complicated in that both the A- and B-distributions are affected, and each involves considering two independently variable infectious periods. Moreover, with measles at any rate, there is little scope for introducing any substantial variations of a into the present model. This can be seen from the fact that if the variance of a were v_a the expected second moment about the origin of the A-distribution would be $2\sigma^2 + 2v_a$, and the observed value only 3·48. However, it can be shown that even if v_a were, say, approximately unity, although the effect on the goodness-of-fit would be very small so far as the A-distribution is concerned, there would be an appreciable flattening of the B-curve. This is clearly a matter that requires further investigation.

6. Allowance for misclassification of chains

So far we have assumed that the basic chains of the mathematical model (almost trivial for families of two) can be correctly identified. For the class with only one case there is no possibility of error. When there are two cases, however, and the A- and B-distributions

overlap, then there is a definite chance of misclassification. In the example discussed in §5, where the two distributions overlap only to a very small extent, the effect is likely to be only a minor one confined to a few borderline observations. With more substantial overlapping it is desirable to make proper allowance for it in the analysis. This can be done by introducing a new parameter, $\xi = 1-\eta$, which is the prior probability that a family is of type A. Maximum-likelihood scoring for the five unknown parameters, λ, m, a, σ and ξ, can then be carried through as before. It is, however, considerably more complicated than the case already discussed. For this reason an indication only of the modified procedure will be given. A full treatment *in extenso* may be worth while elsewhere if data requiring it appear.

First consider the information available from the number of families, C, with one case only. The expected proportion is clearly $\eta e^{-\lambda a}$, and this must now be referred to the whole sample, N, and not merely to $B+C$ as before. The analogue of (2) is thus

$$\binom{N}{C} \eta^C e^{-C\lambda a} (1-\eta e^{-\lambda a})^{N-C}. \tag{26}$$

Next consider the $N-C$ families which are liable to misclassification with regard to the A- and B-distributions. If we write the frequency functions in (1) and (5) as f_1 and f_2, the contribution to the likelihood from any family with two cases separated by an interval of y days is

$$\frac{\xi f_1(y) + \eta(1-e^{-\lambda a}) f_2(y)}{1-\eta e^{-\lambda a}}. \tag{27}$$

Combining the contributions from (26) and (27) gives for the log likelihood

$$L = C\log\eta - C\lambda a + \sum_y \log\{\xi f_1(y) + \eta(1-e^{-\lambda a}) f_2(y)\}, \tag{28}$$

where the summation is taken over the $N-C$ values of y.

Now suppose that there are A and B families that are almost certainly of types A and B, respectively, and D families of uncertain classification. Almost certain classification means that the value of y is such that one of the two frequencies $f_1(y)$ and $f_2(y)$ is negligibly small. In doubtful cases it may be safer to take $A = B = 0$. For the A and B almost certain families we therefore put $f_2 = 0$ and $f_1 = 0$, respectively. This leads to the modified expression

$$\begin{aligned} L = {} & A\log\xi + (B+C)\log\eta - C\lambda a + B\log(1-e^{-\lambda a}) \\ & + \sum_A \log f_1 + \sum_B \log f_2 + \sum_D \log\{\xi f_1 + \eta(1-e^{-\lambda a}) f_2\}. \end{aligned} \tag{29}$$

The expressions (28) or (29) can be used as a basis for the standard technique. Although the scores and information functions are now much more complicated, they contain certain components that are the same as in the simpler situation.

I am particularly indebted to Dr R. E. Hope Simpson of the Cirencester Public Health Laboratory Service for allowing me to make use of his excellent epidemiological records, and also for many stimulating discussions. I should also like to thank Mrs Tamara Hazlewood for carrying out the computations on which the numerical results of this paper are based.

REFERENCES

Bailey, N. T. J. (1953). The use of chain-binomials with a variable chance of infection for the analysis of intra-household epidemics. *Biometrika*, **40**, 279.

Bailey, N. T. J. (1954). A statistical method of estimating the periods of incubation and infection of an infectious disease. *Nature, Lond.*, **174**, 139.

Bailey, N. T. J. (1955). Some problems in the statistical analysis of epidemic data. *J. R. Statist. Soc.* B, **17**, 35.

New York W.P.A. (1942). *Tables of Probability Functions*, vol. **2**.

THE BEHAVIOUR OF AN ESTIMATOR FOR A SIMPLE BIRTH AND DEATH PROCESS

By J. H. DARWIN

Applied Mathematics Laboratory, D.S.I.R., New Zealand

1. Introduction

The simplest birth and death process is one in which there are constant probabilities $\lambda\,dt$ and $\mu\,dt$ of an individual of a population respectively giving birth to a new individual in $(t, t+dt)$ or dying in $(t, t+dt)$. If there are N_0 individuals of the population alive at time 0, the probability generating function (p.g.f.) of the number alive at time τ is

$$\left(\sum_{n=0}^{\infty} p_n z^n\right)^{N_0} = \left[\frac{\mu(\alpha-1)}{\lambda\alpha-\mu} + \sum_{1}^{\infty} \frac{\alpha(\lambda-\mu)^2\lambda^{n-1}(\alpha-1)^{n-1}z^n}{(\lambda\alpha-\mu)^{n+1}}\right]^{N_0}$$

$$= \left[\frac{\mu(\alpha-1)+(\lambda-\mu\alpha)z}{\lambda\alpha-\mu-\lambda(\alpha-1)z}\right]^{N_0}$$

$$= \left[\frac{a+bz}{c-dz}\right]^{N_0}, \tag{1}$$

say, where α is written for $\exp((\lambda-\mu)\tau)$. Thus the probability that the population size is n at time τ is

$$\left(\frac{a}{c}\right)^{N_0} \sum_{r=0}^{\min(n, N_0)} \binom{N_0}{r}\binom{N_0+n-r-1}{n-r}\left(\frac{b}{a}\right)^r\left(\frac{d}{c}\right)^{n-r}.$$

The question of estimating the constants λ and μ has been considered by Anscombe (1953), Kendall (1949, 1952) and Moran (1951, 1953). In his 1949 paper Kendall discusses the case in which it is known that $\mu = 0$, and observations are taken at times $\tau, 2\tau, \ldots, k\tau$. When $\mu = 0$ it is possible to express the probability of a population size n in a much simpler form and find the maximum-likelihood (m.l.) estimate of λ. In the other four papers it is mainly supposed that observation is continued until a certain number of events (births or deaths) has occurred. Such observation may not always be practicable, but observation at regular intervals may be feasible. The complex form of the above probability then almost prohibits the use of m.l. estimation. The estimates of the constants then used must preferably have only a small bias and must be reasonably accurate compared with the m.l. estimates for continuous observation. We discuss the bias and variance of an estimate of α when k evenly spaced population counts are made. If it is also known how many events have occurred by the end of these k counts, the m.l. estimate of μ/λ for continuous observation is available.

2. The estimate of the rate of increase, $\exp((\lambda-\mu)\tau)$

Suppose the population, known to be N_0 at time 0, is counted at times $\tau, 2\tau, \ldots, k\tau$ and that its size at time $i\tau$ is N_i. Then the m.l. estimate of $\alpha \equiv \exp((\lambda-\mu)\tau)$, (*a*) when it is known that $\mu = 0$ (Kendall, 1949) and (*b*) when it is known that $\lambda = 0$ is

$$X_k \equiv \frac{N_1+N_2+\ldots+N_k}{N_0+N_1+\ldots+N_{k-1}}. \tag{2}$$

It is suggested that X_k be used to estimate α when neither λ nor μ is known to be 0. The estimate is intuitively a reasonable one, since the average value of the ratio of corresponding terms in the numerator and denominator is α. X_k is in general biased (see §2·1) and demonstrably inconsistent (see §2·4) as k tends to infinity for a range of values of τ for given λ, μ and N_0. However no simple unbiased estimate of α suggests itself (e.g. $(1/k)\sum_1^k N_i/N_{i-1}$ requires a stopping rule if, for some $j, N_j, N_{j+1}, \ldots, N_k = 0$ and its average value, relative to this rule, will in general be biased). Also X_k has the virtues that its bias (see §2·2) and variance (see §3) tend to zero as N_0 tends to infinity, and that its asymptotic variance then can be made of comparable size to that of the m.l. estimate of α for continuous observation by making observations over a greater period (see §3·1).

2·1 *Bias of* X_k

In finding $E(X_k)$ we may average for X_k first with respect to N_k for a fixed $N_1, \ldots, N_{k-1}$, then average the result of this over N_{k-1} for fixed $N_1, \ldots, N_{k-2}$, and so on.

Thus

$$E(X_k \mid N_1, \ldots, N_{k-1}) = E\left[\frac{N_1 + \ldots + N_{k-2} + N_{k-1}(1+\alpha)}{N_0 + \ldots + N_{k-1}}\right], \tag{3}$$

since

$$E(N_k \mid N_1, \ldots, N_{k-1}) = \alpha N_{k-1}.$$

When averaging over N_{k-1} we observe that the right-hand side is convex from above in N_{k-1}. Then Jensen's convex function inequality for N_{k-1} shows that

$$E(X_k \mid N_1, \ldots, N_{k-2}) \leqslant E\left[\frac{N_1 + \ldots + N_{k-3} + N_{k-2}(1+\alpha+\alpha^2)}{N_0 + \ldots + N_{k-3} + N_{k-2}(1+\alpha)}\right].$$

The same argument repeated till N_0 is reached shows that

$$E(X_k) \leqslant \alpha \equiv \exp((\lambda - \mu)\tau). \tag{4}$$

Since $\log x$ is also a function of x convex from above

$$E(\log x) \leqslant \log E(x) \quad \text{and} \quad E(\log X_k) \leqslant \log E(X_k) \leqslant \log \alpha = (\lambda - \mu)\tau.$$

That is, X_k is a negatively biased estimate of $\alpha \equiv \exp((\lambda - \mu)\tau)$, and $\log X_k$ is a negatively biased estimate of $(\lambda - \mu)\tau$. A closer consideration shows that equality in (4) is never attained for finite N_0 and k and positive τ.

2·2. *N_0 large*

This bias becomes small when N_0 is large. For, suppose x is a positive random variable and A, B, C and D are positive constants. Then

$$\begin{aligned} E_x\left(\frac{A+Bx}{C+Dx}\right) &= E\left[\frac{A+Bx}{C+D\bar{x}}\right] - E\left[\frac{D(A+Bx)(x-\bar{x})}{(C+D\bar{x})^2}\right] + E\left[\frac{D^2(A+Bx)(x-\bar{x})^2}{(C+Dx)(C+D\bar{x})^2}\right] \\ &\geqslant \frac{A+B\bar{x}}{C+D\bar{x}} - \left[\frac{DB\,\mathrm{var}\,(x)}{(C+D\bar{x})^2}\right], \end{aligned} \tag{5}$$

whence

$$E\left[\frac{N_1+\ldots+N_{k-2}+N_{k-1}(1+\alpha)}{N_0+\ldots+N_{k-1}}\,\middle|\,N_1,\ldots,N_{k-2}\right]$$

$$\geqslant \frac{N_1+\ldots+N_{k-3}+N_{k-2}(1+\alpha+\alpha^2)}{N_0+\ldots+N_{k-3}+N_{k-2}(1+\alpha)} - E\left(\frac{(1+\alpha)\,\gamma N_{k-2}}{[N_0+\ldots+N_{k-3}+N_{k-2}(1+\alpha)]^2}\right), \qquad (6)$$

where γN_{k-2} is $\mathrm{var}\,(N_{k-1}\,|\,N_{k-2})$ and in fact $\gamma = [(\lambda+\mu)/(\lambda-\mu)]\,\alpha(\alpha-1)$. If we continue this process on the first term of the right-hand side downward through $N_{k-2},\ldots,N_1$ we achieve a series of bias terms of the same sign. The above one is numerically less than or equal to $(1+\alpha)\,\gamma\alpha^{k-2}/N_0$, and they are all of this order in N_0. In fact the bias is

$$\leqslant (\gamma/N_0)\,[1(1+\alpha)\,\alpha^{k-2}+\ldots+(1+\alpha+\ldots+\alpha^{k-2})\,(1+\alpha+\ldots+\alpha^{k-1})\,1], \qquad (7)$$

and this tends to zero as N_0 tends to infinity, for fixed α and k.

2·3. *Leading term for large* N_0

We use the usual asymptotic value for $E(X_k)$ formed by the average value of the quadratic term in the Taylor expansion of X_k in powers of $(N_1-\overline{N}_1),\ldots,(N_k-\overline{N}_k)$. Suppose X'_k is the value of X_k when each N_i is set equal to $\overline{N}_i$. Then the bias

$$\simeq \sum_{i,j=1}^{k} \tfrac{1}{2}(\partial^2 X'_k/\partial\overline{N}_i\,\partial\overline{N}_j)\,\mathrm{cov}\,(N_i,N_j).$$

The cubic term is of order $1/N_0^2$, and

$$\mathrm{cov}\,(N_i,N_j) = N_0\alpha^j(\alpha^i-1)\,(\lambda+\mu)/(\lambda-\mu) \quad \text{for} \quad j\geqslant i.$$

The bias is, to first order in $1/N_0$,

$$\frac{(\lambda+\mu)\,(\alpha-1)}{N_0(\lambda-\mu)\,(\alpha^k-1)^2}[k\alpha^k-(k-1)\,\alpha^{k+1}-\alpha]$$

$$= -\frac{(\lambda+\mu)\,(\alpha-1)^3}{N_0(\lambda-\mu)\,(\alpha^k-1)^2}[(k-1)\,\alpha^{k-1}+(k-2)\,\alpha^{k-2}+\ldots+\alpha]. \qquad (8)$$

This term disappears when k becomes large for α greater than 1 but remains finite for α less than or equal to 1. This suggests that the behaviour of $\lim_{k\to\infty} E(X_k)$ when N_0 is not necessarily large may vary with the value of α.

2·4. *$E(X_k)$ for large k*

We have been unable to find results for large k holding for the whole range of μ/λ, but those given below indicate that the bias is most serious when λ is less than or equal to μ.

This is to be expected as the probability that N_k is 0 then tends to 1 as k tends to infinity, when τ is fixed. There is thus a high probability for large k that the last ratio of corresponding terms in the numerator and denominator in X_k is 0 which is less than α.

We may divide the process of averaging X_k over the values of $N_1,\ldots,N_k$ into

(*a*) summation over N_k for any fixed $N_1,\ldots,N_{k-1}$. Then

(*b*) summation over all the sets $(N_1,\ldots,N_{k-1})$ of which no member is zero and

(*c*) summation over the $(k-1)$ groups in which the first member to be zero is in turn $N_1,\ldots,N_{k-1}$.

Then $E(X_k)$ has contributions from (b) and (c) after (a) has been done.

(a) First for any fixed $N_1, N_2, \ldots, N_{k-1}$

$$\sum_{N_k} X_k \operatorname{Prob}(N_1, \ldots, N_k) = \frac{N_1 + \ldots + N_{k-2} + N_{k-1}(1+\alpha)}{N_0 + \ldots + N_{k-1}} \operatorname{Prob}(N_1, \ldots, N_{k-1}). \tag{9}$$

(b) The ratio in (9), for any set of positive values of $N_1, \ldots, N_{k-1}$ is less than or equal to $1+\alpha$. The contribution to $E(X_k)$ from (b) is then less than or equal to $(1+\alpha)\left(1 - \sum_{r=1}^{k-1} P_r\right)$ where P_r is the probability that N_r is the first zero $N_i, i = 1, 2, \ldots$. P_r is the total probability that N_r is zero, minus the total probability that N_{r-1} is zero,

$$= \left(\frac{\mu(\alpha^r - 1)}{\lambda\alpha^r - \mu}\right)^{N_0} - \left(\frac{\mu(\alpha^{r-1} - 1)}{\lambda\alpha^{r-1} - \mu}\right)^{N_0}. \tag{10}$$

Then $\sum_1^{k-1} P_r = \left(\frac{\mu(\alpha^{k-1} - 1)}{\lambda\alpha^{k-1} - \mu}\right)^{N_0}$ and the contribution from (b) is less than or equal to

$$(1+\alpha)\left[1 - \left(\frac{\mu(\alpha^{k-1} - 1)}{\lambda\alpha^{k-1} - \mu}\right)^{N_0}\right].$$

(c) If N_r, when r is one of $1, \ldots, k-1$, is the first N_i to be zero

$$X_k = 1 - N_0/(N_0 + \ldots + N_{r-1}) \leqslant 1, \tag{11}$$

and the contribution to $E(X_k)$ from (c) is less than or equal to $\sum_1^{k-1} P_r$.

Hence from (a), (b) and (c),

$$E(X_k) \leqslant 1 + \alpha - \alpha\left(\frac{\mu(\alpha^{k-1} - 1)}{\lambda\alpha^{k-1} - \mu}\right)^{N_0}. \tag{12}$$

Now if λ is greater than μ this upper bound is approximately $1 + \alpha - \alpha(\mu/\lambda)^{N_0}$ for large k. The inequality

$$\alpha > (\lambda/\mu)^{N_0} \tag{13}$$

determines, for given λ, μ and N_0, a value τ_0 of τ such that X_k is inconsistent with respect to k for $\tau \geqslant \tau_0$.

If the term $N_0/(N_0 + \ldots + N_{r-1})$ of (11) which has so far been omitted is taken into account, the upper bound in (12) is reduced by

$$\sum_{r=1}^{k-1} \sum_{N_i} [N_0/(N_0 + \ldots + N_{r-1})] \operatorname{Prob}(N_1, \ldots, N_{r-1}, 0), \tag{14}$$

where the inner summation is taken over the non-zero values of $N_1, \ldots, N_{r-1}$, when N_r is zero.

Consideration of (14) might for some values of λ and μ considerably lower the upper bound of $E(X_k)$ and so decrease the interval τ_0. No general necessary and sufficient condition for the consistency of the estimate X_k as k becomes large has emerged from such a consideration. However, some interesting particular results can be obtained.

If $\lambda = \mu$ that part of (14) which comes from N_1 being zero is $[\lambda\tau/(1+\lambda\tau)]^{N_0}$ and

$$E(X_k) \leqslant 2 - [\lambda\tau/(1+\lambda\tau)]^{N_0} - [(k-1)\lambda\tau/(1+(k-1)\lambda\tau)]^{N_0}$$

$$\simeq 1 - (\lambda\tau/(1+\lambda\tau))^{N_0} \quad \text{for large } k. \tag{15}$$

Hence there is always inconsistency no matter what finite value N_0 has.

If λ is less than μ a similar use of the first term of (14) shows that there is always inconsistency if $N_0 = 1$. But for N_0 greater than 1 such use involves conditions on λ and μ for

inconsistency to be demonstrable. If, however, only the second term of (14) is included we may prove a stronger result. This second term is

$$\sum_{N_1}[N_0/(N_0+N_1)]\left(\frac{\mu(1-\alpha)}{\mu-\lambda\alpha}\right)^{N_1}\text{Prob}(N_1),$$

where the summation is over non-zero values of N_1 and $\text{Prob}(N_1)$ is that given by the coefficient of z^{N_1} in the p.g.f. (1). Then by Schwarz's inequality the term has a value

$$\geqslant N_0 P_2^2/[\sum_{N_1}(N_0+N_1)[\mu(1-\alpha)/(\mu-\lambda\alpha)]^{N_1}\text{Prob}(N_1)]$$

$$\geqslant P_2^2/(P_2+\alpha). \tag{16}$$

Hence
$$E(X_k)\leqslant 1-P_1-P_2^2/(P_2+\alpha)+\alpha\left[1-\left(\frac{\mu(1-\alpha^{k-1})}{\mu-\lambda\alpha^{k-1}}\right)^{N_0}\right]$$

$$\doteq 1-P_1-P_2^2/(P_2+\alpha) \tag{17}$$

for large k, and any α less than 1.

Suppose α is small. Then this lower bound

$$\doteq \alpha N_0(\mu-\lambda)/(N_0(\mu-\lambda)+\mu).$$

It is always possible therefore to find a τ, say τ_1, making α small enough for this lower bound to be less than α. Hence X_k is inconsistent with respect to k for τ greater than τ_1.

2·41. *The case* $\mu = 0$

That consistency with respect to k is possible is shown by the limiting behaviour of $E(X_k)$ when it is known that $\mu = 0$ and X_k is used to estimate $\exp(\lambda\tau)$. Then the joint p.g.f. of $N_0, N_1, \ldots, N_k$ can be written down, and hence that of $N_k - N_0$, and $N_0 + \ldots + N_{k-1}$. Suppose this is $\phi(u, v)$, where u and v are carriers for $N_k - N_0$ and $N_0 + \ldots + N_{k-1}$, respectively. Then

$$E(X_k) = 1 + \int_0^1 \frac{\partial}{\partial u}[\phi(u,v)]_{u=1}\,dv/v. \tag{18}$$

It follows from (1) with $\mu = 0$ that

$$\phi(u,v) = \left[\frac{v^k}{\alpha^k-(\alpha-1)\,u(\alpha^{k-1}+\alpha^{k-2}v+\ldots+v^{k-1})}\right]^{N_0}. \tag{19}$$

Then

$$E(X_k) = 1+(\alpha-1)\int_0^1 \frac{N_0 v^{N_0 k-1}(\alpha^k-v^k)\,(\alpha-v)^{N_0}}{(\alpha^k(1-v)+(\alpha-1)\,v^k)^{N_0+1}}\,dv$$

$$\geqslant 1+(\alpha-1)\int_0^1 \frac{N_0 v^{N_0 k-1}(\alpha^k-1)\,(\alpha-1)^{N_0}}{(\alpha^k(1-v)+\alpha-1)^{N_0+1}}\,dv$$

$$= 1+(\alpha-1)^{N_0+1}(1-\alpha^{-k})\int_0^1 v^{N_0 k-1}\,d(\alpha^k(1-v)+\alpha-1)^{-N_0}$$

$$= 1+(\alpha-1)(1-\alpha^{-k})-(\alpha-1)^{N_0+1}(1-\alpha^{-k})(N_0 k-1)\int_0^1 v^{N_0 k-2}(\alpha^k(1-v)+\alpha-1)^{-N_0}\,dv. \tag{20}$$

The integral here is positive and less than $\int_0^1(\alpha^k(1-v)+\alpha-1)^{-N_0}dv$ which can be evaluated. For N_0 greater than 1 it is $O(\alpha^{-k})$ and for $N_0 = 1$ it is $O(k\alpha^{-k})$. Hence the right-hand side has α for its limit as k tends to infinity. But $E(X_k)$ is less than or equal to α. Hence $\lim_{k\to\infty} E(X_k)$ exists and is α.

2·42. *The case* $\lambda = 0$

The actual limiting bias can be found when it is known that $\lambda = 0$, so that X_k is an estimate of $\exp(-\mu\tau)$. By the same method as was used in § 2·41 we find that

$$E(X_k) \to 1 - N_0(1-\alpha)^{N_0} \int_0^1 \frac{v^{N_0-1}}{(1-\alpha v)^{N_0}} dv. \tag{21}$$

For example for $N_0 = 1$, when the bias is biggest,

$$E(X_k) \to 1 - ((1-\alpha)/\alpha) \log [1/(1-\alpha)]. \tag{22}$$

This takes values 0, 0·107, 0·234, 0·389, 0·598, 0·744, for $\alpha = 0$, 0·2, 0·4, 0·6, 0·8, 0·9, respectively.

3. The variance of X_k for large N_0

§ 2·4 shows that there is not in general an improvement in the average value of X_k as k becomes large, but § 2·2 shows that a large N_0 means that the bias is small. It is thus natural to consider the accuracy of X_k rather for large N_0 than for large k. We shall use the variance of X_k for large N_0 as a yardstick for measuring the efficiency of this type of regular observation compared with continuous observation.

The limiting form of the variance of X_k to order $1/N_0$ is

$$\sum_{i,j=1}^{k} \frac{\partial X_k'}{\partial \bar{N}_i} \frac{\partial X_k'}{\partial \bar{N}_j} \operatorname{cov}(N_i, N_j) = \frac{(\lambda+\mu)\alpha(\alpha-1)^2}{N_0(\lambda-\mu)(\alpha^k-1)}. \tag{23}$$

The coefficient of $1/N_0$ is small for large k when α is greater than or equal to 1, but for α less than 1, increasing k gives little improvement.

We wish to compare this variance with that of the m.l. estimate of α when observation is continuous. The justification for the use of the m.l. estimate is that if N_0 is large there is effectively a large number of independent replicates. Suppose such observation is made over a period t and that r events have happened at intervals $\tau_1, \tau_2, \ldots, \tau_r$ so that

$$\tau_1 + \tau_2 + \ldots + \tau_r \leqslant t.$$

Then the likelihood of this happening when there were N_0 members at the beginning of observation is

$$\exp[-N_0(\lambda+\mu)\tau_1] N_0 \nu_0 \exp[-n_1(\lambda+\mu)\tau_2] n_1 \nu_1 \ldots \exp[-n_{r-1}(\lambda+\mu)\tau_r] n_{r-1}\nu_{r-1} \times \exp[-n_r(\lambda+\mu)(t-(\tau_1+\ldots+\tau_r))]. \tag{24}$$

In this ν_i is either λ or μ depending on whether the $(i+1)$th event that has happened is a birth or a death; n_{i+1} is n_i+1 if ν_i is λ and is n_i-1 if ν_i is μ. Here $n_0 \equiv N_0$. Then the log likelihood is

$$L = -(\lambda+\mu)X + \sum_0^{r-1} \log n_i + B\log\lambda + D\log\mu, \tag{25}$$

where

$$X = n_0\tau_1 + n_1\tau_2 + \ldots + n_{r-1}\tau_r + n_r(t-(\tau_1+\ldots+\tau_r)),$$

B is the number of births, and D the number of deaths in time t. The m.l. estimate of $\lambda-\mu$ is $(B-D)/X$. The variance of the m.l. estimate, $\exp(t(B-D)/X)$ of $\alpha_t \equiv \exp[(\lambda-\mu)t]$, is, to the first order in $1/N_0$,

$$t^2\alpha_t^2\left[\frac{(\bar{B}-\bar{D})^2}{\bar{X}^4}\operatorname{var} X - 2\frac{\bar{B}-\bar{D}}{\bar{X}^3}\operatorname{cov}(X, B-D) + \frac{1}{\bar{X}^2}\operatorname{var}(B-D)\right], \tag{26}$$

where $\bar{B}-\bar{D}$ and $\bar{X}$ are mean values.

Now $$B-D = n_t - N_0,$$

$$E(B-D) = N_0(\alpha_t - 1),$$

and $$\operatorname{var}(B-D) = N_0(\lambda+\mu)\alpha_t(\alpha_t-1)/(\lambda-\mu). \tag{27}$$

Again, since
$$\left.\begin{aligned} X &= \int_0^t n_u\,du, \\ E(X) &= N_0(\alpha_t-1)/(\lambda-\mu), \\ E(X^2) &= \int_0^t\int_0^t E(n_u n_v)\,du\,dv = 2\int_0^t du\int_0^u E(n_v^2)\,e^{(\lambda-\mu)(u-v)}\,dv, \text{ etc.}, \\ \text{and}\quad E(X(B-D)) &= \int_0^t E(n_u(n_t-N_0))\,du \\ &= \int_0^t E(n_u^2)\,e^{(\lambda-\mu)(t-u)}\,du - \bar{X}N_0, \text{ etc.} \end{aligned}\right\} \tag{28}$$

Then for large N_0 $$\operatorname{var}\left[\exp\left(\frac{(B-D)\,t}{X}\right)\right] \simeq \frac{(\lambda+\mu)\,\alpha_t^2\,(\log\alpha_t)^2}{N_0(\lambda-\mu)\,(\alpha_t-1)}. \tag{29}$$

We may first compare (29) with (23) when k is 1 and t is taken as τ. Since this is a comparison of the variance of the m.l. estimate of α for continuous observation over τ with the variance of X_k for one observation at τ we may define the ratio of these variances as the efficiency of X_1. This ratio is

$$\left[\frac{\frac{1}{2}(\lambda-\mu)\,\tau}{\sinh\left(\frac{1}{2}(\lambda-\mu)\,\tau\right)}\right]^2. \tag{30}$$

This is the same formula for $\lambda-\mu$ as was Kendall's formula for λ (Kendall, 1949).

3·1. *Further comparison of efficiencies*

Suppose a population is known to have a large size N_0 at time 0. Then we may compare the accuracy of the following estimates of $\alpha \equiv \exp((\lambda-\mu)\,\tau)$.

(*a*) X_k when the population is counted only at $\tau, 2\tau, \ldots, k\tau$, and

(*b*) $\exp((B-D)\,\tau/X)$ when the population is continuously observed over the interval $(0, t)$.

The limiting variance of X_k for large N_0 is (23). That of $\exp((B-D)\,\tau/X)$ is (29) multiplied by $(\tau/t)^2\,\alpha^{2-2t/\tau}$. Suppose $t/\tau = u$. This latter variance is then

$$\frac{(\lambda+\mu)\,\alpha^2(\log\alpha)^2}{N_0(\lambda-\mu)\,(\alpha^u-1)}.$$

Hence the accuracy of the estimates (*a*) and (*b*) for large N_0 is the same if

$$\alpha^u - 1 = (\alpha^k-1)\,\alpha(\log\alpha)^2/(\alpha-1)^2. \tag{31}$$

For a given α and k a unique solution of this for u always exists. For, the left-hand side is always monotonic in u for a given α.

If α is greater than 1 the left-hand side goes from zero to infinity as u goes from zero to infinity, and this range includes the value of the right-hand side.

If α tends to 1, u tends to k.

If α is less than 1 the left-hand side goes from zero to -1 as u goes from zero to infinity. By comparing the derivatives of $\log\alpha$ and $(\alpha-1)/\sqrt{\alpha}$ we can show that $\alpha(\log\alpha)^2/(\alpha-1)^2$ is always less than 1 for all positive α other than 1. Hence the range zero to -1 includes the value of the right-hand side.

It also follows that, except when $\alpha = 1$, k is always greater than u. That is, $k\tau$ is greater than t, as indeed is intuitively obvious as (a) can in general only be as efficient as (b) if X_k embodies some information not available to (b), i.e., if at least the last of the k observations is made after t.

When α approaches 1 the rate of change in population size becomes small and continuous information becomes of little extra assistance in its estimation.

Table 1 gives the values of u for $k = 1$ (when (23) is exact) and a range of values of α. Then

$$u = \frac{\log[1+\alpha(\log\alpha)^2/(\alpha-1)]}{\log\alpha}. \tag{32}$$

We thus have the following result.

We know that a population has a large size N_0 at time zero and wish to make an estimate of its rate of increase when it develops according to a simple birth and death process. The most natural estimate to take is the m.l. estimate formed, as in § 3·0, from the record of continuous observations of the population over a period t.

Table 1

α	u	α	u
0·1	0·386	2·0	0·971
0·2	0·649	3·0	0·941
0·4	0·890	4·0	0·916
0·6	0·972	5·0	0·898
0·8	0·955	10·0	0·838
1·0	1·000		

This section and § 2·2 show that the estimate X_1 formed from a single population count at time τ is approximately unbiased, and of as low a variance as the m.l. estimate derived from continuous observation over $(0, t)$, if τ is greater than a particular value, t/u, greater than t. For a wide range of α Table 1 shows that this value is not very much greater than t. For example, for α in (0·294, 16·1) it is less than $1{\cdot}25t$.

4. Estimation of μ/λ

If $N_1, \ldots, N_k$ are the only observed quantities, estimation of μ/λ is likely to be very inaccurate since the range of values of μ/λ giving the same set $N_1, \ldots, N_k$ with a reasonable probability is very large. The extra information of the actual number of births and deaths that have occurred may be available. Then, since the m.l. estimate of μ/λ is, from (25), D/B, this estimate can be used, and it will be fully efficient for large N_0. The theoretical difficulty is that a rule would have to be devised to cope with the situations $B = 0$, or D and $B = 0$. In the first case $\lambda = 0$ is a suitable inference. The second case should of course be avoided

by N_0 being made large enough. The probability of there being no events at all in time τ is $\exp[-N_0(\lambda+\mu)\tau]$ and of there being no births $[\mu/(\lambda+\mu)+(\lambda/(\lambda+\mu))\exp(-(\lambda+\mu)\tau)]^{N_0}$, both of which can be made small for non-zero λ and μ by an increase in N_0.

I am greatly indebted to the referee for his suggestions for the improvement of this paper.

REFERENCES

ANSCOMBE, F. J. (1953). *J. R. Statist. Soc.* B, **15**, 1–29.
KENDALL, D. G. (1949). *J. R. Statist. Soc.* B, **11**, 230–64.
KENDALL, D. G. (1952). *Ann. Inst. Poincaré*, **13**, 43–108.
MORAN, P. A. P. (1951). *J. R. Statist. Soc.* B, **13**, 141–6.
MORAN, P. A. P. (1953). *J. R. Statist. Soc.* B, **15**, 241–5.

EXAMINATION OF A QUANTUM HYPOTHESIS BASED ON A SINGLE SET OF DATA

By S. R. BROADBENT

The British Coal Utilization Research Association

1. Introduction

In an earlier paper (Broadbent, 1955) the following situation was discussed: an experimenter makes the observations $y_1, y_2, \ldots, y_n$ and wishes to compare his data with the *quantum hypothesis*

$$y_i = \beta + 2r_i\delta + \epsilon_i \quad (i = 1, \ldots, n). \tag{1}$$

Here β and 2δ are constants (2δ is the *quantum*), r_i is zero or an integer and ϵ_i is the error of observation. On this hypothesis the data will be grouped about regularly spaced means and the grouping will be apparent if the ϵ_i are small in comparison with δ. The alternative hypothesis, that the y_i are not so grouped, but are distributed unimodally or rectangularly, is called the rectangular hypothesis.

Solutions were suggested in the previous paper to the following problems:

(i) estimation of β and of 2δ,

(ii) estimation of the variance of ϵ_i,

(iii) testing whether to accept a quantum hypothesis which is independent of the data used in the test.

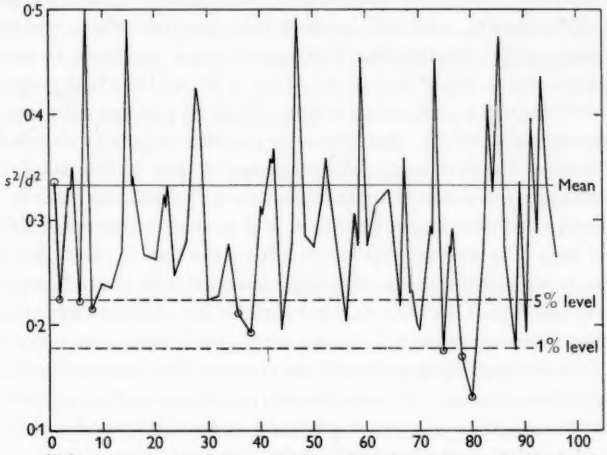
Sometimes the quantum hypothesis to be tested is not independent of the data alleged to support it, but the data have actually suggested the hypothesis. In effect the experimenter says to the statistician: 'My data appears to be grouped about regularly spaced means. I have no independent evidence of the positions of these groups, and the body of scientific knowledge (and my prior belief) is neither strongly for nor strongly against such a quantum hypothesis. Is the apparent grouping a coincidence, or does it indicate some physical reality?' This question was specifically excluded in the previous paper, and is the subject of the present investigation.

The situation may be clarified by two analogies which show that this type of problem is common in statistics. The first is with the χ^2 test of goodness of fit. Here the agreement between data and hypothesis is measured by χ^2; when the hypothesis is independent of the data the number of degrees of freedom used in the test equals the number of classes used minus one. But if the data are used to estimate parameters which the hypothesis does not specify, the χ^2 test must be modified. The modification given by Fisher (1924) consists simply in reducing the number of degrees of freedom used in the test by the number of parameters estimated. If the data were used for other purposes, for example, to suggest whether one or another type of distribution be fitted to the data or what the class limits should be, the modification would no longer be simple. The second analogy is with the analysis of time series, which is closely related to this problem. In periodogram analysis agreement with the trial periods may be measured by the intensity S^2. If the hypothesis (i.e. the trial period) is independent of the data, the significance of S^2 may be tested. But this independence is often not the case; for example, Kendall (1946) showed that of

Beveridge's 18 or 19 'real' periods in the analysis of the wheat-price index, at least three-quarters were spurious. Kendall concludes that tests of significance in the periodogram remain undiscovered; Irwin (1956) and Rudra (1955) have recently discussed such problems.

We shall for simplicity suppose $\beta = 0$ in the hypothesis (1), and we write $2d$ for a quantum considered after examination of the data whereas 2δ is the quantum in a hypothesis independent of the data. The statistic s^2/d^2 (see the previous paper and (2) below) measures the agreement between the data and a proposed quantum $2d$.

On the rectangular hypothesis, s^2/δ^2 has mean $\frac{1}{3}$, variance $\frac{4}{45}n$ and is approximately normally distributed. A significantly small value of s^2/δ^2 indicates the validity of the quantum hypothesis. The value of s^2/d^2 may have to be considerably smaller than a conventional significance point of s^2/δ^2 to validate the quantum hypothesis, since use of the data to suggest a quantum implies a low value of s^2/d^2, on either hypothesis. How small s^2/d^2 may have to be is indicated by the experiment described below.



s^2/d^2 as a function of $1/d$ for a single set of 20 observations uniformly distributed

Fig. 1

Consider the variation in s^2/d^2 for a fixed set of observations as d changes. We restrict attention to values of d smaller than the largest observation, since the data cannot give information about larger quanta. It is convenient to consider $1/d$ increasing in value from the reciprocal of the largest observation. Then s^2/d^2 appears as a violently oscillating continuous function. An example is given in Fig. 1, which shows the stationary values of s^2/d^2 (joined by straight lines) for a set of twenty observations, nineteen of which were randomly drawn from a uniform distribution between 0 and 1, 1 being taken as the twentieth observation. This sampling procedure is discussed in § 3. On this figure are shown also the mean and the lower 5 and 1 % significance points of s^2/δ^2 for twenty observations on the rectangular hypothesis (taken from Table 3 of Broadbent, 1955). Although the observations actually obeyed the rectangular hypothesis, s^2/d^2 fell below the 5 % point in no less than thirteen intervals for $1/d$ between 1 and 100, and below the 1 % point in three intervals. At $1/d = 80{\cdot}1$, s^2/d^2 nearly reached the 0·1 % point which is 0·1273. At all these values of $1/d$, the data would have been conventionally judged to be significantly grouped.

The experimenter should realize that s^2/d^2 may be made as small as he pleases by suitable choice of d; indeed, if the data are rational, s^2/d^2 vanishes for infinitely many values of d. When he considers possible quanta, he is in effect looking for the minima of s^2/d^2 as a function of $1/d$. Small values of s^2/d^2 at large values of $1/d$ are expected even on the rectangular hypothesis. We proceed on the assumption that the experimenter is looking for minima of s^2/d^2 at small values of $1/d$ (i.e. large values of the quantum $2d$).

To find the distribution of such minima on the rectangular hypothesis seems to be a very difficult problem in analysis. Conventional analysis of extreme-value problems does not appear to apply to the minima of such functions as s^2/d^2, in which the random element (the observations) enters once and for all, and is thereafter treated as fixed. The distribution is accordingly investigated by sampling methods in the third section, the calculations required being developed in the second section. The conclusions may be of assistance in testing proposed quanta for significance.

Two bases for this test should be emphasized. It is assumed that the experimenter is searching for possible quanta, and will propose the quantum which seems to him most unlikely on the rectangular hypothesis. This search for a quantum is analogous to the estimation of a parameter in the χ^2 test of goodness of fit, and the test proposed is in effect analogous to the well-known modification to this χ^2 test. We do not take into consideration the likely but imponderable event, that the experimenter originally decided to search for a quantum only because his particular data suggested it (just as the usual modification to the χ^2 test does not take into account a possible choice of hypothesis tested). Secondly, the experimenter's method of choosing a quantum will probably not agree exactly with the method described here. The first consideration affects the test in the following way: if our test shows that the experimenter's data are not inconsistent with the rectangular hypothesis, we may confidently assert that the data does not support his quantum hypothesis. Similarly, if the experimenter does not assume $\beta = 0$ but estimates a value for β, and his s^2/d^2 is not inconsistent with the rectangular hypothesis, we may use the test described below to reject his quantum hypothesis. The second consideration operates in the opposite way, that is, even if our test would reject his proposed quantum, it remains possible that he could with improved methods propose a quantum which fits his data more closely. It is, however, likely that intuitive estimators of quanta do not give estimates very different from optimum estimators; see, for example, the discussion on Prof. Thom's data (1955). In this case the onus is still on the experimenter to propose the quantum which fits the data better.

2. Analysis of observations

Suppose $y_1, y_2, \ldots, y_n$ are the (positive) observations made by the experimenter and arranged in increasing order of magnitude; the adjustments necessary if two or more of the y_i are equal are trivial and not detailed here. The agreement of these observations with any proposed quantum $2d$ is clearly measured by

$$s^2/d^2 = \sum_{i=1}^{n} (y_i - 2r_i d)^2/(nd^2). \tag{2}$$

Here r_i is zero or that integer which minimizes $|\, y_i - 2r_i d \,|$; this does not define r_i uniquely when $y_i = (2m+1)\,d$ for m zero or an integer, and in this case we take $r_i = m+1$. The set $\{r_1, r_2, \ldots, r_n\}$ we write $\mathbf{r}(d)$; the vector is a function of the observations and of d.

We noticed in the first section that the largest value of d we need consider is $d = y_n$. Suppose that values smaller than d^* need not be considered; for example, d^* might be the order of precision with which the observations were made. We consider the behaviour of $\mathbf{r}(d)$ and s^2/d^2 for a fixed set of observations as d varies from y_n to d^*.

The ith element of $\mathbf{r}(d)$ increases from m to $(m+1)$ as d attains the value $y_i/(2m+1) = d_{im}$. We arrange the values of d_{im} $(i = 1, 2, \ldots, n;\ m = 0, 1, \ldots)$ in decreasing order of magnitude from $d_{n0} = y_n$. The interval between d_{im} and next smallest element of the ordered set we call the interval $[d_{im}]$; the interval is closed at d_{im} and open at the other end. We notice that at $d = y_n$ the vector $\mathbf{r}(d)$ is $\{0, 0, \ldots, 1\}$; the vector remains constant for all values of d within any interval $[d_{im}]$, and changes in the way just described, only at the values d_{im}.

It is obvious that s^2/d^2 is a differentiable function of d within any interval $[d_{im}]$. At d_{im} it may be shown that s^2/d^2 is continuous but not differentiable; the term $(y_i - 2r_i d)/nd^2$ has slopes $\pm 2/(ny_i)$ at $d_{im} \mp \epsilon$ as $\epsilon \to 0$ (independent of m), and so has an upward-pointing cusp there. It follows that s^2/d^2 cannot have a stationary value or minimum at any d_{im}, but may have a maximum there. All minima of s^2/d^2 therefore occur strictly within the intervals $[d_{im}]$ and may be found by equating to zero the differential of s^2/d^2 with regard to d, in which $\mathbf{r}(d)$ is treated as constant.

The necessary condition for a minimum is

$$2d = \sum_{i=1}^{n} y_i^2 \Big/ \sum_{i=1}^{n} r_i y_i, \tag{3}$$

and here s^2/d^2 has the value

$$4\left\{\sum_{i=1}^{n} r_i^2 \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} r_i y_i\right)^2\right\} \Big/ \left\{n \sum_{i=1}^{n} y_i^2\right\}. \tag{4}$$

It follows that in each interval $[d_{im}]$ there is at most one stationary value. In the interval $\mathbf{r}(d)$ is constant, and insertion of the corresponding values of r_i in (3) gives a value of d which may or may not lie in the interval $[d_{im}]$. If d does lie in the interval, the stationary value of s^2/d^2 is given by (4); if it does not, there is no stationary value in the interval.

It is now possible to lay down a procedure which gives a list of all $2d$ we might suspect to be quanta, down to any predetermined small value $2d^*$. The procedure is lengthy, but is suitable for use with mechanical methods of computing; the calculations described below were carried out on Hollerith machines. The procedure is as follows:

(*a*) Compute $d_{im} = y_i/(2m+1)$ for $i = 1, 2, \ldots, n$ and $m = 0, 1, \ldots$. The calculation stops for each i at the first d_{im} less than d^*.

(*b*) Arrange the d_{im} computed in (*a*) for all i and m in decreasing order of magnitude, down to the first d_{im} less than d^*.

(*c*) Compute $\sum_{i=1}^{n} y_i^2$.

(*d*) Compute $\sum_{i=1}^{n} r_i y_i$ in each interval $[d_{im}]$. In the first interval this sum is just y_n; its value in the interval d_{im} is y_m larger than its value in the preceding interval. Its values are therefore easily found in succession.

(*e*) Compute $\sum_{i=1}^{n} r_i^2$ in each interval $[d_{im}]$. In the first interval the sum is 1, and its value in the interval $[d_{im}]$ is $(2m+1)$ larger than its value in the preceding interval

(*f*) Compute in each interval $[d_{im}]$, from (*c*), (*d*) and (3), the value of d at which s^2/d^2 is stationary, if such a value exists. If the computed value lies in the interval, it is 'accepted'.

(*g*) If the value of d obtained in (*f*) is accepted, compute the corresponding s^2/d^2 from (*c*), (*d*), (*e*) and (4).

These calculations result in a list of stationary values of s^2/d^2, and the values of d at which they occur. Maxima and points of inflexion will occur in the list but can be neglected; the cusps which occur at d_{im} do not appear in the list. If, of two minima, that corresponding to the smaller d is at the larger (or equal) value of s^2/d^2, it may generally be ignored. This is because small values of s^2/d^2 are expected at small values of d even on the rectangular hypothesis. The list of minima may therefore generally be reduced to a *list of successive absolute minima*, in order of decreasing d.

A short example of this calculation will now be given. Suppose five values are observed, 1, 9, 10, 19 and 21, and that 3 is the value taken for d^*. Once the d_{im} are computed and ordered, the y_i and $(2m+1)$ appropriate to each interval $[d_{im}]$ may be given as in Table 1.

Table 1. *Calculation of possible estimates for a quantum*

d_{im}	y_i	$(2m+1)$	$\sum_{i=1}^{n} r_i y_i$	$\sum_{i=1}^{n} r_i^2$	d	s^2/d^2
21						
	21	1	21	1	(23·43)	—
19						
	19	1	40	2	12·30	0·294
10						
	10	1	50	3	9·84	0·367
9						
	9	1	59	4	8·33	0·370
7						
	21	3	80	7	(6·15)	—
6·3						
	19	3	99	10	4·97	0·032
4·2						
	21	5	120	15	(4·10)	—
3·8						
	19	5	139	20	3·54	0·292
3·3						
	10	3	149	23	3·30	0·350
3						

The running totals of the second and third columns appear as the fourth and fifth columns. The estimate of d given by (3) follows, the values not accepted (not lying within their appropriate intervals) being given in brackets. Finally, the corresponding value of s^2/d^2 is given.

The list of successive absolute minima of s^2/d^2 has in this case only two entries, 0·294 and 0·032. Of the five entries in the last column of Table 1, four are near the expected value $\frac{1}{3}$ on the rectangular hypothesis. But at $d = 4{\cdot}97$ the low value of 0·032 has been obtained for s^2/d^2 which indicates that the observations are grouped about zero or integer multiples of $2d$, i.e. 0, 9·94 and 19·88. In this simple case, a value of $2d$ near 10 could have been deduced by inspection. An experimenter who made such observations would be liable to announce a quantum law, and to account for the small discrepancies from the law as experimental error. An estimator of the variance about the means was given in the previous paper (Table 2, Broadbent, 1955); for these data we estimate the standard deviation to be 0·9.

3. s^2/d^2 ON THE RECTANGULAR HYPOTHESIS

We return to the question asked by the experimenter in the first section, and now phrase it: 'By studying my observations I have found a possible quantum $2d$ such that s^2/d^2 seems to me remarkably small for such a small $1/d$. Is its value so small as to cast doubt on the rectangular hypothesis?'

We attempt to answer this question by comparing the experimenter's values of s^2/d^2 and $1/d$ with values found in a sampling experiment in which the rectangular hypothesis held good. Two difficulties arise in making this comparison.

The first difficulty is that points in the $(s^2/d^2, 1/d)$ plane cannot at present be completely ordered in their departure from the expected value on the rectangular hypothesis. Suppose the experimenter obtains the small values s_1^2/d_1^2 and s_2^2/d_2^2 at $1/d_1$ and $1/d_2$ and that $1/d_1 < 1/d_2$. If $s_1^2/d_1^2 \leqslant s_2^2/d_2^2$, he will certainly consider $2d_1$ a more likely value for a quantum than $2d_2$ for the reasons already given. But if $s_1^2/d_1^2 > s_2^2/d_2^2$, he must choose on rather intuitive grounds which is the more likely quantum. For example, in Fig. 1, the successive absolute minima of s^2/d^2 have been circled, and at each value below a conventional significance level the corresponding $2d$ might be chosen as quantum. In other words, each of the successive absolute minima (circled points in Fig. 1) corresponds to a possible candidate for consideration as a quantum. Is the ordinate at $1/d = 2{\cdot}0$ (where $s^2/d^2 = 0{\cdot}22$) more startling than the ordinate at the much larger $1/d = 80{\cdot}1$ (where s^2/d^2 has the smaller value $0{\cdot}13$)? In the absence of knowledge of the distribution of such minima, and without adopting some principle of ordering, the question cannot be answered, and for this reason an exact test of significance of the rectangular hypothesis does not seem possible. However, a general comparison of the experimenter's values with those obtained on the rectangular hypothesis may still be made.

The second difficulty is that the exact distribution of successive minima on the rectangular hypothesis is not known, and the sampling experiment was limited in two ways. It was possible to take only a few values of n; this disadvantage is to some extent overcome later. Only a rectangular parent distribution has been used. Although the distributions of s^2/δ^2 for a unimodal and a rectangular parent distribution are very similar, it may not be true to say the same for the distributions of successive minima of s^2/d^2. In the comparison suggested below, the largest observation is an important statistic, and its distribution for the two types of parent is very different. The comparison described is therefore only valid when a rectangular parent distribution is a reasonable alternative to the quantum hypothesis, although it may be illuminating in other cases.

The sampling experiment was carried out by the Mathematics Division of the National Physical Laboratory. For a given value of n, $(n-1)$ numbers between 0 and 1 were taken from the Rand random numbers, and the number 1 added to the set. In this way sets were obtained with the same distribution as that given by the following procedure: take n random numbers distributed uniformly between 0 and a $(0 < a < \infty)$ and divide each by the largest number taken. At first sight the two procedures appear different, but in each the ith ordered observation y_i may be shown to have the distribution

$$(n-1)\binom{n-2}{i-1} y_i^{i-1}(1-y_i)^{n-i-1}\,dy_i \quad (0 \leqslant y_i \leqslant 1, i = 1, 2, \ldots, n-1),$$

and similarly the joint distribution of the y_i are the same in each case. Since s^2/d^2 is dimensionless the actual value of a and the fact that the largest observation in the calculations is

always 1 are irrelevant in applications. That is, the distribution of s^2/d^2 obtained in the sampling experiment is directly comparable with that of s^2/d^2 obtained when the parent distribution is uniform between 0 and a.

In the sampling experiment 100 sets for $n = 5$ and for $n = 20$ observations were produced, and 50 sets for $n = 10$ and for $n = 50$. For each set, the calculation described in § 2 was completed and a table of successive absolute minima of s^2/d^2 was formed. The small value d^* was 0·01, i.e. possible quanta as small as 1/50 of the range of the observations were considered. The size of the calculations is indicated by the number of d_{im} formed: 140,000.

For one set of each sample size a complete table of accepted d and corresponding s^2/d^2 was printed. One of these tables of stationary values of s^2/d^2 is given in Fig. 1 as a graph of s^2/d^2 against $1/d$. The other three were of a similar nature. Four comments may be made:

(i) For each sample size n, the density of stationary values seems to remain approximately constant as $1/d$ increases.

(ii) The density of stationary values increases with sample size n.

(iii) The range of the oscillations of s^2/d^2 decreases as n increases.

(iv) Cusps (possibly maxima) at d_{im} have been ignored.

The tables of successive minima have about the same average number of entries (5·5), i.e. there are nine successive minima in Fig. 1, but in the average sample only 5·5. The scatter diagrams in Figs. 2–5 show these entries as points in the $(s^2/d^2, 1/d)$ plane. As $1/d$ increases, the values of the minima fall sharply at first and after that decrease slowly. Fig. 2 shows that for $n = 5$ very small values of s^2/d^2 are obtained even for small values of $1/d$. The points circled in Fig. 2 are described below. It is therefore difficult (as might be expected) to substantiate a quantum hypothesis with as few as five observations; startling agreement with the hypothesis may be obtained, but if the experimenter has been free to select the quantum similar agreement may be obtained on the rectangular hypothesis.

On the rectangular hypothesis, $\sqrt{n}\{\frac{1}{3} - s^2/\delta^2\}$ is approximately normally distributed, with mean zero and variance 4/45; it is independent of n. Inspection of Figs. 3–5 suggests the following conjecture, that $\sqrt{n}\{\frac{1}{3} - s^2/d^2\}$ is approximately independent of n and so depends on d only, provided that n is greater than 5. On this conjecture, we may use this statistic in a test of a quantum hypothesis for any n greater than 5, its distribution being obtained from the data of the sampling experiment. A large positive value of $\sqrt{n}\{\frac{1}{3} - s^2/d^2\}$ indicates that the observations are grouped about zero or integer multiples of $2d$. Fig. 6 gives successive maxima of $\sqrt{n}\{\frac{1}{3} - s^2/d^2\}$ for 50 sets of $n = 10$, 20 and 50 observations (i.e. half the data of Fig. 4, and all Figs. 3 and 5 are represented in Fig. 6).

Fig. 6 supports the conjecture in that the points for the three values of n appear quite well mixed. For $1/d > 50$ there are 45, 45 and 50 points for $n = 10$, 20 and 50 respectively. The means of the respective ordinates are 0·69 (S.D. 0·028), 0·76 (S.D. 0·028) and 0·76 (S.D. 0·017). The largest difference between the three pairs of means is greater than zero by 1·8 times its S.D.

The test of a quantum hypothesis proposed consists in comparing the values $\sqrt{n}\{\frac{1}{3} - s^2/d^2\}$ and y_n/d obtained by the experimenter with Fig. 6 (y_n being equal to 1 in the sampling experiment). If the experimenter's point appears consistent with the other points on the figure, there is no reason to doubt the rectangular hypothesis. If his point lies above and to the left of the others, it indicates that his quantum hypothesis is obeyed by the data more precisely than chance warrants on the rectangular hypothesis. As a rough guide it may be said that

$$\sqrt{n}\{\tfrac{1}{3} - s^2/d^2\} > 1$$

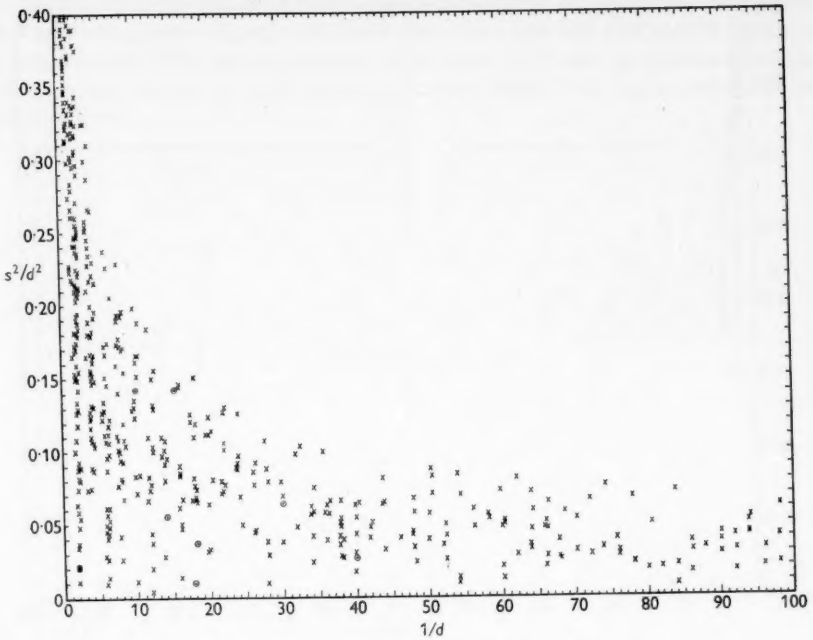


Fig. 2. Successive minima of s^2/d^2: 100 sets for $n=5$.

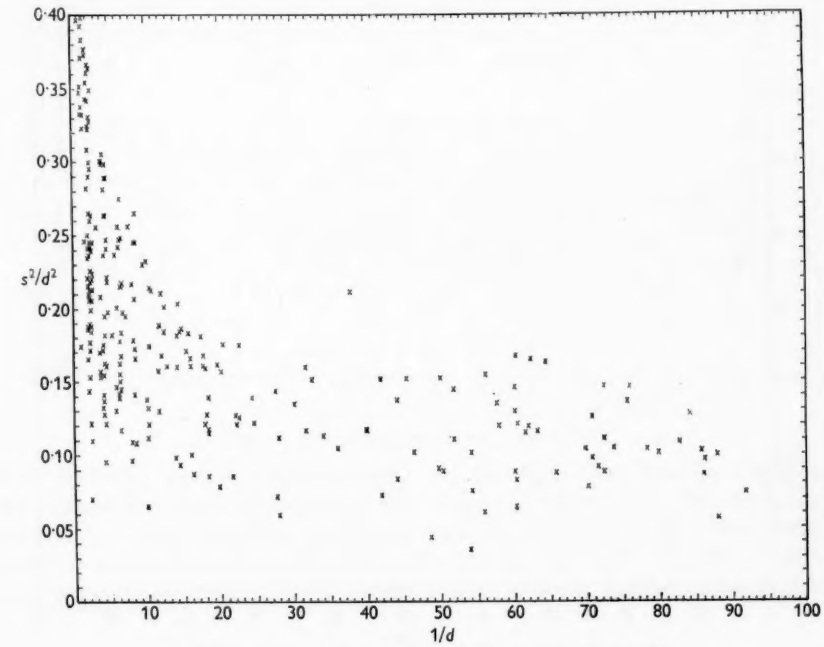


Fig. 3. Successive minima of s^2/d^2: 50 sets for $n=10$.

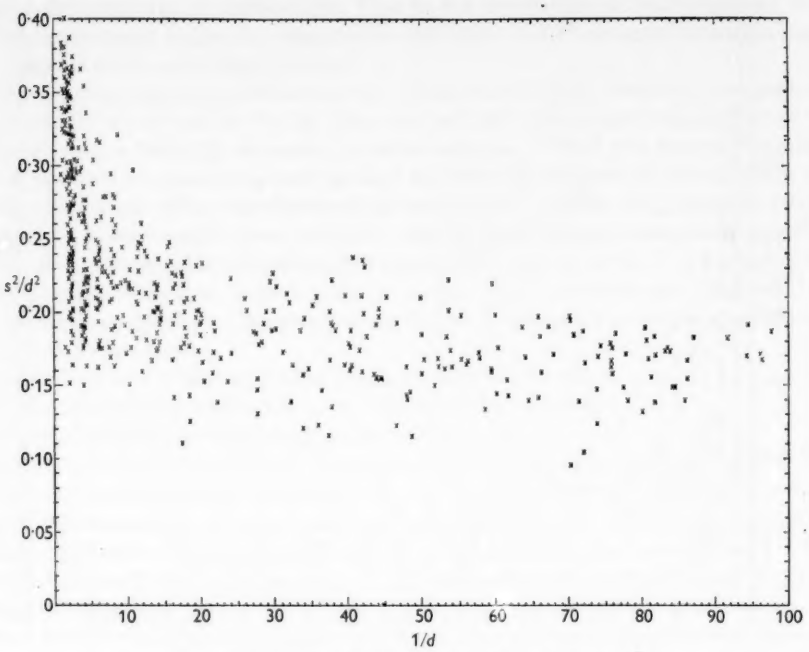


Fig. 4. Successive minima of s^2/d^2: 100 sets for $n = 20$.

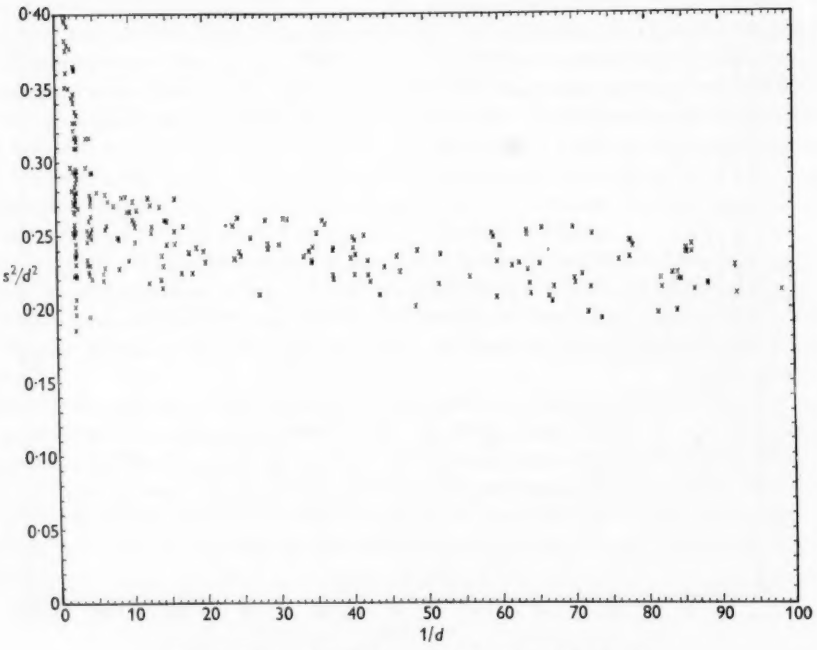


Fig. 5. Successive minima of s^2/d^2: 50 sets for $n = 50$.

is the condition that the data must fulfil to validate the quantum hypothesis. In comparison, the 5, 1 and 0·1 % points of $\sqrt{n}\{\frac{1}{3} - s^2/\delta^2\}$ are about 0·49, 0·69 and 0·92, since its mean is zero and variance 4/45. When the experimenter is at liberty to choose the quantum, in order to validate his hypothesis, he must obtain agreement better than the conventional one in a thousand level.

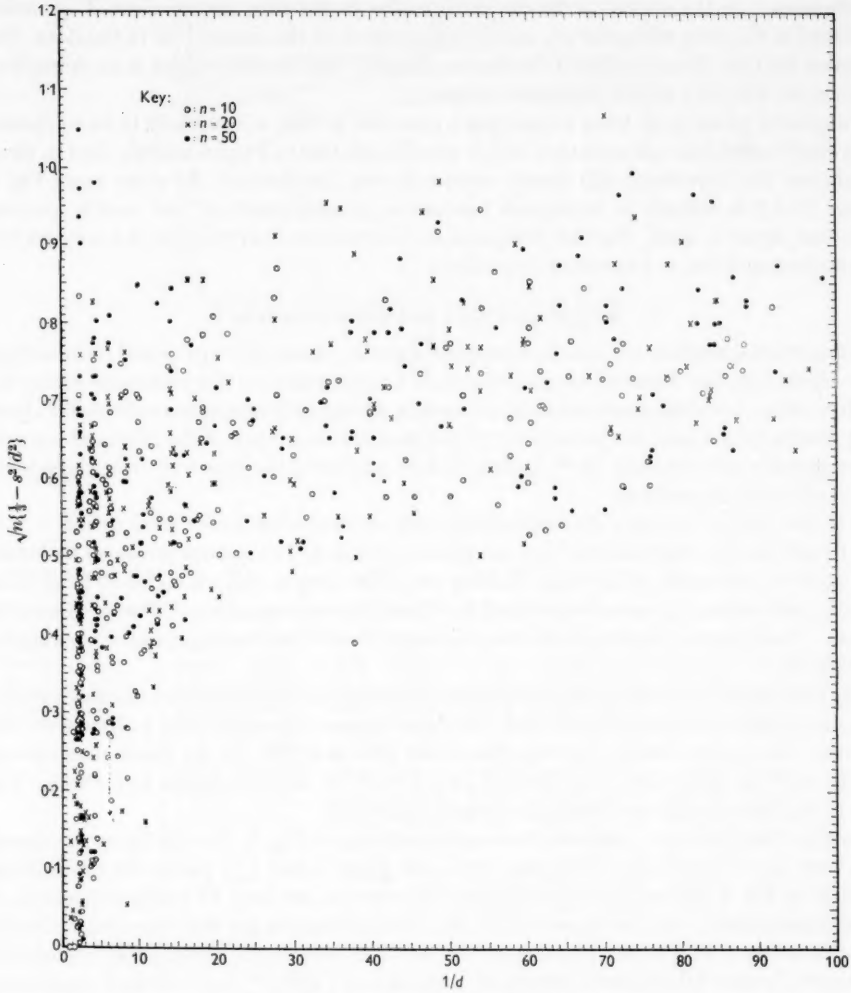


Fig. 6. Successive maxima of $\sqrt{n}\{\frac{1}{3} - s^2/d^2\}$.

It will probably occur to the reader that a quantum hypothesis based on a single set of data could also be tested by a random division of the data into two sets. The first half of the data would be used to estimate a possible quantum, and this value would be tested with the second half of the data. Two difficulties arise in this treatment. The first is that two statisticians using this method might disagree in their judgement of the same data (because

it had been divided in different ways), although of course their probabilities of error of the first and second kinds are the same. The second difficulty is that the decision, arrived at after inspecting all the data, to test for the existence of a quantum will entail a rough guess at the value of the quantum. This guess will generally influence the estimation of the quantum from half the data, for the estimation generally requires knowledge of the coefficients r_i or the ability to decide which modes in the data are spurious. A quantum deduced in this way will generally not be independent of the second half of the data. The division by Prof. Thom of Druid Circles into English and Scottish circles is an example of the second difficulty of this treatment of data.

In general terms, even when a quantum hypothesis is true, it is unlikely to be confirmed by a single set of data unless the S.D. of ϵ_i is small in relation to δ and n is large. And in these conditions the hypothesis will hardly require formal justification. In other cases Fig. 6 shows that it is difficult to distinguish between a genuine quantum 'law' and a spurious quantum fitted to data. Further independent observations may then be the only way of strengthening belief in a quantum hypothesis.

4. An example of a quantum hypothesis

After surveying the data of excitation energies of nuclei, Grant (1952) proposed the quantum hypothesis that the observed energy levels were integer multiples of a particular energy for each nucleus. Accurate measurements of two or more energy levels were available for about fifty nuclei; Grant gave the probability of the random occurrence of the observed integral sequences as approximately 10^{-30}, in other words he believed the data were overwhelmingly in favour of the hypothesis.

The five largest numbers of observations made on individual nuclei were with $n = 7$, 8, 10, 10 and 12; the observations† (y_i) are given in Table 2. The quanta proposed by Grant are based on inspection of the data. Making use of the vectors $\mathbf{r}(d)$ which follow from these quanta, estimates of $2d$ were recalculated by (3) and the corresponding s^2/d^2 were calculated by (4). It will now be considered whether the data of Table 2 tend to disprove the rectangular hypothesis.

We first notice that the agreement between columns y_i and $2r_i d$ for each nucleus appears intuitively good, and the values of s^2/d^2 calculated appear to be well below $\frac{1}{3}$. However, two of these values are certainly not significant (for $^{82}_{36}$Kr and $^{15}_{7}$N), i.e. for these observations $\sqrt{n}\{\frac{1}{3} - s^2/d^2\}$ is below the 5 % level of $\sqrt{n}\{\frac{1}{3} - s^2/\delta^2\}$ on the rectangular hypothesis. The remaining three would uncritically be judged significant.

In Fig. 7 the data are represented on a graph similar to Fig. 6. On this figure are drawn also very approximate curves for the mean and upper 5 and 1 % points for the minima plotted on Fig. 6 (for reasons already given these curves can only be rough indications of significance levels). It now appears that the two agreements for the two nuclei already mentioned are in fact rather worse than one would expect on the rectangular hypothesis, two more ($^{17}_{8}$O and $^{11}_{5}$B) appear to be typical, and one only ($^{214}_{84}$Ra C′) can be judged significant.

Similar conclusions can be drawn from the following comparison. In the last row of Table 2 are given the reported standard deviations of the observations, and in the row above the estimated standard deviations about the quanta calculated. For the first four nuclei, the reported standard deviation is too small to account for the variation actually found. For $^{214}_{84}$Ra C′ the two are equal.

† Data privately communicated to the author.

For reasons already given, the remainder of the data is not sufficiently numerous to throw doubt on the rectangular hypothesis. For example, the seven sets of $n = 5$ observations gave the points circled in Fig. 2 on the $(s^2/d^2, 1/d)$ diagram. Although some low values of s^2/d^2 were obtained, it was hardly possible to get values lower than those found on the rectangular hypothesis.

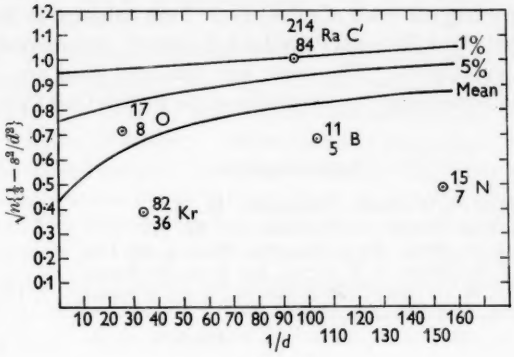


Fig. 7

Table 2. *Energy levels of five nuclei (Grant's data)*

Nucleus ...	$^{82}_{36}$Kr		$^{15}_{7}$N		$^{11}_{5}$B		$^{17}_{8}$O		$^{214}_{84}$RaC′	
No. of observations ...	7		8		10		10		12	
	y_i	$2r_id$	y_i	$2r_id$	y_i	$2r_id$	y_i	$2r_id$	y_i	$2r_id$
Observed and fitted energy levels (MeV.)	0·550	0·542	5·276	5·338	2·141	2·144	0·875	0·635	0·608	0·612
	0·610	0·619	5·305	5·338	4·457	4·467	3·02	3·176	1·283	1·286
	0·689	0·697	6·328	6·322	5·033	5·003	3·88	3·811	1·412	1·409
	0·768	0·774	7·164	7·164	6·752	6·789	4·54	4·446	1·663	1·353
	0·825	0·851	7·309	7·305	6·802	6·789	5·17	5·081	1·844	1·837
	1·038	1·006	8·315	8·288	7·297	7·325	5·72	5·716	2·015	2·021
	1·315	1·316	9·156	9·131	8·565	8·576	6·33	6·352	2·138	2·143
	—	—	10·816	10·817	8·921	8·934	6·93	6·987	2·268	2·266
	—	—	—	—	9·185	9·112	7·60	7·622	2·439	2·450
	—	—	—	—	9·272	9·291	8·23	8·257	2·513	2·511
	—	—	—	—	—	—	—	—	2·697	2·695
	—	—	—	—	—	—	—	—	2·880	2·878
$2d$	0·07739		0·14048		0·17867		0·63515		0·06124	
s^2/d^2	0·1864		0·1612		0·1169		0·1072		0·0331	
$\sqrt{n\{\frac{1}{3}-s^2/d^2\}}$	0·3887		0·4867		0·6844		0·7194		1·0401	
y_n/d	34·0		154·0		103·7		25·9		94·1	
Estimate of σ	0·017		0·029		0·031		0·104		0·006	
Reported σ	0·003		0·006		0·008		0·020		0·006	

By itself the nucleus $^{214}_{84}$Ra C′ might be considered to validate a quantum hypothesis, but as a whole that data must be thought of as consistent with the rectangular hypothesis.

Part of this paper was written at the Atomic Energy Research Establishment, Harwell, during a vacation consultancy, and I am indebted to the Establishment, particularly to Dr J. Howlett, for supporting the work of calculation. I am indebted to Mr C. W. Nott, of the Mathematics Division of the National Physical Laboratory, who carried out the calculations on the sampling experiment. The paper has benefited from my discussions with Prof. G. A. Barnard, and is published by permission of the British Coal Utilization Research Association.

REFERENCES

Broadbent, S. R. (1955). *Biometrika*, **42**, 45.
Fisher, R. A. (1924). *J. R. Statist. Soc.* **87**, 442.
Grant, P. J. (1952). *Proc. Phys. Soc. Lond.* A, **65**, 150.
Irwin, J. O. (1956). *J. R. Statist. Soc.* B (in the Press).
Kendall, M. G. (1946). *The Advanced Theory of Statistics*, **2**, 435.
Rudra, A. (1955). *Sankhyā*, **15**, 9.
Thom, A. (1955). *J. R. Statist. Soc.* A, **118**, 275.

THE NUMBER OF NEW SPECIES, AND THE INCREASE IN POPULATION COVERAGE, WHEN A SAMPLE IS INCREASED

By I. J. GOOD and G. H. TOULMIN

A sample of size N is drawn at random from a population of animals of various species. Methods are given for estimating, knowing only the contents of this sample, the number of species which will be represented r times in a second sample of size λN; these also enable us to estimate the number of different species and the proportion of the whole population represented in the second sample. A formula is found for the variance of the estimate; when $\lambda > 2$, this variance becomes in general very large, so that the estimate is useless without some modification. This difficulty can be partly overcome, at least for $\lambda < 5$, by using Euler's method with a suitable parameter or the methods described by Shanks (1955) to hasten the convergence of the series by which the estimate is expressed. The methods are applied to samples of words from *Our Mutual Friend*, to an entomological sample, and to a sample of nouns from Macaulay's essay on Bacon.

1. Introduction

We present here a further development of the theory expounded by Good (1953); that paper will be referred to, for brevity, by the letter G throughout.

We imagine a random sample of size N, the *basic sample*, to be drawn from an infinite population of animals of various species, and suppose that n_r distinct species are each represented exactly r times in the sample, so that

$$\sum_{r=1}^{\infty} r n_r = N. \tag{1}$$

We write

$$d = \sum_{r=1}^{\infty} n_r,$$

the total number of distinct species in the sample. It is convenient (though, as was pointed out in G, not essential) to suppose that the total number of distinct species in the population is a known finite number s, so that we can calculate

$$n_0 = s - d, \tag{2}$$

the number of species not represented in the sample. If the actual value of s is not known, all our results will remain true if it is arbitrarily assumed to be any sufficiently large number. As in G (p. 237), the larger n_1 is, the more applicable our results are. In G it was shown that certain properties of the population could be deduced approximately from the sample frequencies n_r; in particular, the total *coverage* of the sample (i.e. the proportion of the population represented in the sample, which is the sum of the population frequencies p_μ of the species represented) is approximately

$$1 - \frac{\mathscr{E}(n_1)}{N} \simeq 1 - \frac{n_1}{N}, \tag{3}$$

rovided n_1 is large (G, formula (9)).†

† $\mathscr{E}(n_1)$ is the expected value of the random variable n_1 when our basic sample of N specimens is taken at random. We shall use the same symbol n_1 both for this random variable and for a particular value of it.

We now contemplate taking a *second sample*, of size λN. We describe this as the 'second sample', even though it may be (and in practice probably will be) an enlargement of the basic sample; in this case, of course, $\lambda \geqslant 1$. If the second sample is not an enlargement of the basic sample, it will be termed *independent*; this word may be interpreted in its probabilistic sense, provided that the true statistical hypothesis specifying the population frequencies is momentarily regarded as 'given'. Except in §4 our results apply to both enlargements and independent second samples.

We may now wish, for example:

(*a*) To find the expected coverage of the second sample.

(*b*) To find the expected number of distinct species in the second sample.

(*c*) To find (roughly) the variances of estimates of population parameters which might be made from the second sample.

(*d*) To estimate the term $\mathscr{E}_{2N}(n_{2r} \mid H)$ in formula (22) of G for the variance of n_r.

Results of this type may enable us to decide whether it is worth enlarging our sample, and to what extent, depending on the purposes for which it is required.

For example, consider a teacher of languages who wishes to base his teaching on the population frequencies of words. He will wish to estimate what size of vocabulary should be learnt by a student in order to decrease the need for reference to a dictionary below a certain frequency. It was shown in G how a sample can be used to make such an estimate. The present paper shows in what way the sample can also be used in order to help the decision of whether to carry out more sampling. For instance, in example (iii) of §6 below, the 2048 words of the basic sample had an expected coverage of 87·3 %, and we find that if the sample size were doubled, then the same expected coverage could be obtained by selecting only 1780 words. More work by the teacher means less for the student.

Similarly, an entomologist will often want to know whether to increase a basic sample, and will be able to base his decision largely on the expected number of new species that will be provided by a given amount of sampling. Example (ii) of §6 is an instance of such an application.

Let $n_r(\lambda)$ be the random variable whose value is the number of distinct species represented exactly r times in the second sample.†

We first consider a method that may appeal to statisticians who are accustomed to fit distributions by the method of moments. The method will not, however, be used in our examples if only because of the enormous amount of calculation that it requires. We begin by stating a lemma that is presumably well known, although we cannot give a reference.

LEMMA. (Determination of a set of numbers whose 'factorial moments' are specified.) If

$$b_i = \sum_{r=0}^{\infty} r^{(i)} a_r \quad (i = 0, 1, 2, \ldots), \tag{4}$$

then

$$a_r = \frac{1}{r!} \sum_{i=0}^{\infty} \frac{(-1)^i b_{r+i}}{i!}, \tag{5}$$

at any rate if $a_r = 0$ for all sufficiently large r.‡ Problems (*a*), (*b*) and (*d*) above reduce to the estimation of the numbers $\mathscr{E}(n_r(\lambda))$ for certain values of r when values of $n_r(1) = n_r$ are

† It is convenient here to depart slightly from the notation of G. The number which we write as $\mathscr{E}(n_r(\lambda))$ would there have been denoted by $\mathscr{E}_{\lambda N}(n_r)$. Note that in the case of an enlarged sample, $n_r(\lambda)$ is to be considered as varying as the whole enlarged sample is varied at random, not merely the $(\lambda - 1)\,N$ additional specimens, so that this correspondence of notation still holds.

‡ For a proof under more general conditions, see the Appendix p. 62 below.

observed; the results are given in equations (22), (23) and (32), respectively. The same is true of problem (*c*) when the variances concerned can be expressed in terms of the $n_r(\lambda)$; thus equations (30A) and (31) of G show that this is the case for Yule's 'characteristic', which is an estimate of $\sum_{\mu=1}^{s} p_\mu^2$.

Now
$$\frac{1}{N^{(i)}} \sum_{r=0}^{\infty} r^{(i)} n_r$$
is an unbiased estimate of $c_i = \sum_{\mu=1}^{s} p_\mu^i$ (see, for example, G, p. 245). Similarly

$$\mathscr{E}\left\{\frac{1}{(\lambda N)^{(i)}} \sum_{r=0}^{\infty} r^{(i)} n_r(\lambda)\right\} = \sum_{\mu=1}^{s} p_\mu^i. \tag{6}$$

It may therefore seem reasonable to assume

$$\sum_{r=0}^{\infty} r^{(i)} \mathscr{E}(n_r(\lambda)) \simeq \frac{(\lambda N)^{(i)}}{N^{(i)}} \sum_{r=0}^{\infty} r^{(i)} n_r \left[\simeq \lambda^i \sum_{r=0}^{\infty} r^{(i)} n_r\right], \tag{7}$$

and then to solve for $\mathscr{E}(n_r(\lambda))$ by using the lemma.

In spite of the theoretical interest of this method it seems likely that it is not really adequate. For it depends too much on the estimates of the higher population moments, c_i, and these estimates are subject to large sampling errors. We have therefore not investigated any numerical examples. Instead of proceeding via the factorial moments, we find directly the following relation between $\mathscr{E}(n_r(\lambda))$ and the $\mathscr{E}(n_r)$:

$$\mathscr{E}(n_r(\lambda)) \simeq \lambda^r \sum_{i=0}^{\infty} (-1)^i \binom{r+i}{r} (\lambda-1)^i \mathscr{E}(n_{r+i}) \tag{8}$$

for any integer $r \geqslant 0$ (§2, equation (16)). If we assume that

$$\mathscr{E}(n_{r+i}) \simeq n_{r+i} \tag{9}$$

or
$$\mathscr{E}(n_{r+i}) \simeq n'_{r+i}, \tag{10}$$

where the numbers $n'_1, n'_2, n'_3, \ldots$ are obtained by smoothing the numbers $n_1, n_2, n_3, \ldots$ (see G, §§3, 7, 8), we can estimate the values of $\mathscr{E}(n_r(\lambda))$.

We point out here that the series (8) is not really infinite: for

$$\mathscr{E}(n_{r+i}) = 0 \quad \text{whenever} \quad r+i > N, \tag{11}$$

and the upper limit of summation could therefore be replaced by $N-r$. If $\lambda > 2$, a practical difficulty arises, in that the factor $(\lambda-1)^i$ increases rapidly with i, and so attaches great weight to terms for which $\mathscr{E}(n_{r+i})$ is small and therefore is liable to a large percentage error when estimated from the basic sample. It seems to be practicable to overcome this difficulty, at least for moderate values of λ (say $\lambda \leqslant 5$), by using a summation method to make the series (8) converge rapidly; it is shown in §5 that Euler's method with a suitably chosen parameter q is convenient. We have not, however, been able to justify this procedure by finding a useful error term for the partial sums of the new series obtained.

We mention here two possibilities which we have not investigated practically. First, it may be possible to reach larger values of λ in two (or more) stages: e.g. to estimate $\mathscr{E}(n_1(4))$ we might, instead of using (8) directly with $\lambda = 4$, first estimate $\mathscr{E}(n_1(2))$, $\mathscr{E}(n_2(2))$,

$\mathscr{E}(n_3(2)), \ldots$, then smooth the values obtained, and again apply (8) with $\lambda = 2$ to estimate $\mathscr{E}(n_1(4))$. Secondly, (8) with $1/\lambda$ in place of λ and n_r, $n_r(\lambda)$ interchanged might be used as a check on the results obtained, and this might provide a new method of smoothing.

We are much indebted to Dr J. Wishart for several suggestions and corrections.

2. Estimation of $\mathscr{E}(n_r(\lambda))$

Let $p_\mu (\mu = 1, 2, \ldots, s)$ be the population frequencies of the s species. As in G, equation (10),

$$\mathscr{E}(n_r) = \sum_{\mu=1}^{s} \binom{N}{r} p_\mu^r (1-p_\mu)^{N-r}. \tag{12}$$

(In G, the left-hand side is written as $\mathscr{E}_N(n_r | H)$. As explained above, we use the symbol n_r only with reference to the basic sample, of size N; we omit the H, which refers to the hypothesis that the population frequencies are $\{p_\mu\}$, because we shall not be concerned with expectations on any other hypothesis.) For the second sample, we have similarly, assuming $p_\mu < \frac{1}{2}$ for all μ,

$$\begin{aligned}
\mathscr{E}(n_r(\lambda)) &= \sum_{\mu=1}^{s} \binom{\lambda N}{r} p_\mu^r (1-p_\mu)^{\lambda N - r} \\
&= \sum_{\mu=1}^{s} \binom{\lambda N}{r} p_\mu^r (1-p_\mu)^{N-r} \left(1 + \frac{p_\mu}{1-p_\mu}\right)^{-(\lambda-1)N} \\
&= \sum_{\mu=1}^{s} \binom{\lambda N}{r} p_\mu^r (1-p_\mu)^{N-r} \sum_{i=0}^{\infty} \binom{-(\lambda-1)N}{i} p_\mu^i (1-p_\mu)^{-i} \qquad (13) \\
&= \sum_{i=0}^{\infty} \binom{\lambda N}{r} \binom{-(\lambda-1)N}{i} \sum_{\mu=1}^{s} p_\mu^{r+i} (1-p_\mu)^{N-(r+i)} \\
&= \sum_{i=0}^{\infty} \frac{\binom{\lambda N}{r}\binom{-(\lambda-1)N}{i}}{\binom{N}{r+i}} \mathscr{E}(n_{r+i}). \qquad (14)
\end{aligned}$$

(14) is not rigorously correct, since for $r+i > N$, $\binom{N}{r+i}$ and $\mathscr{E}(n_{r+i})$ both vanish, and the corresponding terms of the series are indeterminate. We notice, however, that if the infinite upper limit for i is replaced by an odd [even] integer, the left-hand side of (13) is greater [less] than the right-hand side†, and the same therefore holds of (14). Thus the partial sums of (14) are alternately greater and less than the left-hand side, and in all practical examples a sufficiently good approximation is reached while $(r+i)$ is still small compared to N. Provided that we use only terms of the series for which $r+i \ll N$ and $i \ll (\lambda-1)N$, we can write

$$\begin{aligned}
\frac{\binom{\lambda N}{r}\binom{-(\lambda-1)N}{i}}{\binom{N}{r+i}} &\simeq \frac{(\lambda N)^r (-(\lambda-1)N)^i (r+i)!}{r!\, i!\, N^{r+i}} \\
&= (-1)^i \lambda^r (\lambda-1)^i \binom{r+i}{r}. \qquad (15)
\end{aligned}$$

Hence

$$\mathscr{E}(n_r(\lambda)) \simeq \lambda^r \sum_{i=0}^{\infty} (-1)^i \binom{r+i}{r} (\lambda-1)^i \mathscr{E}(n_{r+i}), \tag{16}$$

the partial sums erring alternately in excess and defect.

† This follows from the nth Mean Value Theorem applied to $(1+x)^{-(\lambda-1)N}$.

(16) can also be obtained directly by using the Poisson approximation

$$\mathscr{E}(n_r) \doteq \sum_{\mu=1}^{s} e^{-Np_\mu} \frac{(Np_\mu)^r}{r!}. \tag{17}$$

We define an estimate of $\mathscr{E}(n_r(\lambda))$ by

$$\hat{n}_r(\lambda) = \lambda^r \sum_{i=0}^{\infty} (-1)^i \binom{r+i}{r} (\lambda-1)^i n_{r+i}; \tag{18}$$

then (16) gives us $\mathscr{E}(\hat{n}_r(\lambda)) \doteq \mathscr{E}(n_r(\lambda))$.

For the case $r = 0$, we may seem to need to assume the value of s, but this assumption is not really required since we can write

$$\hat{d}(\lambda) = d - \sum_{i=1}^{\infty} (-1)^i (\lambda-1)^i n_i = s - \hat{n}_0(\lambda), \tag{19}$$

so that $\mathscr{E}(\hat{d}(\lambda)) \doteq \mathscr{E}(d(\lambda))$.

We have thus obtained (approximately) unbiased estimates of $\mathscr{E}(n_r(\lambda))$ and $\mathscr{E}(d(\lambda))$ in terms of the observed numbers n_r. We shall almost certainly obtain more accurate estimates if we replace the n_r by smoothed values n'_r; for methods of smoothing see G, §§3, 7, 8.

In the case when the second sample is an enlargement of the basic one, and it is desired to predict the value of $n_r(\lambda)$ *given the basic sample* rather than $\mathscr{E}(n_r(\lambda))$, which we have defined to be the expectation when the whole of the second sample is varied at random, it seems intuitively clear that n_r should not be replaced by n'_r in (18), at any rate when λ is not large, though the later terms probably should be smoothed; for instance, consider the case $\lambda = 1$. We have not attempted a rigorous treatment of this question. The advantage of smoothing is not as great when using formulae like (18) as when using such formulae as G (2′):

$$r^* \doteq (r+1)\, n'_{r+1}/n'_r,$$

involving a ratio of the n'_r. Sometimes a ratio is involved surreptitiously, as in G (6), (6′).

An important point in the argument leading to (14) was the expression of $(1-p_\mu)$ as

$$\left(1+\frac{p_\mu}{1-p_\mu}\right)^{-1},$$

so that expansion led to terms expressible as functions of the $\mathscr{E}(n_r)$. This device can also be used to avoid the approximation made in G (lines 8 and 9 of p. 241) of replacing the expected value of n_{r+m} for a sample of size $N+m$ by its expected value for a sample of size N. The result obtained is (using the notation of G)

$$\mathscr{E}(q_r^m \mid H) = \frac{(r+m)^{(m)}}{(N-r)^{(m)}} \frac{1}{\mathscr{E}(n_r)} \sum_{i=0}^{\infty} \frac{(r+m+i)^{(i)}}{(N-r-m)^{(i)}} \binom{-m}{i} \mathscr{E}(n_{r+m+i}) \tag{20}$$

$$= \frac{(r+m)^{(m)}}{(N-r)^{(m)}} \frac{\mathscr{E}(n_{r+m}) - \dfrac{(r+m+1)\,m}{N-r-m} \mathscr{E}(n_{r+m+1}) + \dots}{\mathscr{E}(n_r)}. \tag{21}$$

The approximations made in G are thus equivalent to replacing $(N-r)^{(m)}$ by N^m and neglecting the terms of the sum after the first; they are reasonable provided $mr \ll N$.

As noted in §1, we may be particularly interested in the coverage of the second sample, and the number of different species it will contain. By (3) and (16) with $r = 1$, the expected coverage is approximately

$$1-\frac{\mathscr{E}(n_1(\lambda))}{\lambda N} \doteq 1-\frac{1}{N}\sum_{i=0}^{\infty}(-1)^i(i+1)(\lambda-1)^i\mathscr{E}(n_{i+1})$$

$$\doteq 1-\frac{1}{N}[n_1-2(\lambda-1)n_2+3(\lambda-1)^2n_3-\ldots] \tag{22}$$

(or, more accurately, the same formula with n_r' in place of n_r).† The expected number of distinct species represented is (by (19)) approximately

$$d+(\lambda-1)n_1-(\lambda-1)^2n_2+\ldots; \tag{23}$$

i.e. in the case of an enlarged sample, the number of new species expected is approximately

$$(\lambda-1)n_1-(\lambda-1)^2n_2+\ldots. \tag{24}$$

Evidently $n_1, n_2, \ldots$ may be replaced by smoothed values in (23) and (24), but d should be replaced by the smoothed value

$$d' = n_1'+n_2'+\ldots$$

only in the case of an independent second sample. Note that (23) and (24) can be proved directly without assuming s to be finite.

3. Variance of the estimates $\hat{n}_r(\lambda)$

In this section we find an expression for the variance of the estimate $\hat{n}_r(\lambda)$ of $\mathscr{E}(n_r(\lambda))$ defined by (18). This must not be confused with the variance of $n_r(\lambda)$, which can be found from the formulae given in G, §5. $\hat{n}_r(\lambda)$ is a linear function of the random variables n_{r+i}, and varies accordingly when we take different basic samples; we can find its variance if we know the variances and covariances of the n_r. We therefore start by calculating these. By the method of G, §5, we find

$$\mathscr{E}(n_rn_s) = \delta_{rs}\mathscr{E}(n_r)+\frac{N!}{r!\,s!\,(N-r-s)!}\sum_{\mu,\nu}^{\mu+\nu}p_\mu^rp_\nu^s(1-p_\mu-p_\nu)^{N-r-s}$$

$$= \delta_{rs}\mathscr{E}(n_r)+\frac{N!}{r!\,s!\,(N-r-s)!}\Big[\sum_\mu\sum_\nu p_\mu^rp_\nu^s(1-p_\mu-p_\nu)^{N-r-s}-\sum_\mu p_\mu^{r+s}(1-2p_\mu)^{N-r-s}\Big], \tag{25}$$

where $\delta_{rr} = 1$, $\delta_{rs} = 0$ if $r \neq s$.

Now

$$(1-p_\mu-p_\nu)^{N-r-s} = \left[(1-p_\mu)(1-p_\nu)\left(1-\frac{p_\mu}{1-p_\mu}\frac{p_\nu}{1-p_\nu}\right)\right]^{N-r-s}$$

$$= (1-p_\mu)^{N-r}\left(1+\frac{p_\mu}{1-p_\mu}\right)^s(1-p_\nu)^{N-s}\left(1+\frac{p_\nu}{1-p_\nu}\right)^r\left(1-\frac{p_\mu}{1-p_\mu}\frac{p_\nu}{1-p_\nu}\right)^{N-r-s}$$

$$= \sum_{i=0}^{s}\binom{s}{i}p_\mu^i(1-p_\mu)^{N-r-i}\sum_{j=0}^{r}\binom{r}{j}p_\nu^j(1-p_\nu)^{N-s-j}$$

$$\times\sum_{k=0}^{N-r-s}(-1)^k\binom{N-r-s}{k}p_\mu^k(1-p_\mu)^{-k}p_\nu^k(1-p_\nu)^{-k}; \tag{26}$$

† It is correct to replace n_1 by n_1', even when the second sample is an enlargement of the basic one, because the more accurate formula for the coverage, G (9′), uses n_1' in place of n_1, so we are interested in $\mathscr{E}(n_1(\lambda))$ rather than the expected value of $n_1(\lambda)$ given the basic sample.

and
$$(1-2p_\mu)^{N-r-s} = \left[(1-p_\mu)\left(1-\frac{p_\mu}{1-p_\mu}\right)\right]^{N-r-s}$$
$$= \sum_{i=0}^{N-r-s}(-1)^i\binom{N-r-s}{i}p_\mu^i(1-p_\mu)^{N-r-s-i}. \tag{27}$$

Substituting from (26) and (27) in (25),

$$\mathscr{E}(n_r n_s) = \delta_{rs}\mathscr{E}(n_r) + \frac{N!}{r!\,s!\,(N-r-s)!}\Bigg[\sum_{i,j,k}(-1)^k\binom{s}{i}\binom{r}{j}\binom{N-r-s}{k}$$
$$\times\left(\sum_\mu p_\mu^{r+i+k}(1-p_\mu)^{N-r-i-k}\right)\left(\sum_\nu p_\nu^{s+j+k}(1-p_\nu)^{N-s-j-k}\right)$$
$$-\sum_i(-1)^i\binom{N-r-s}{i}\sum_\mu p_\mu^{r+s+i}(1-p_\mu)^{N-r-s-i}\Bigg]$$
$$= \delta_{rs}\mathscr{E}(n_r) + \frac{N!}{r!\,s!\,(N-r-s)!}\Bigg[\sum_{i,j,k}(-1)^k\frac{\binom{s}{i}\binom{r}{j}\binom{N-r-s}{k}}{\binom{N}{r+i+k}\binom{N}{s+j+k}}\mathscr{E}(n_{r+i+k})\mathscr{E}(n_{s+j+k})$$
$$-\sum_i(-1)^i\frac{\binom{N-r-s}{i}}{\binom{N}{r+s+i}}\mathscr{E}(n_{r+s+i})\Bigg]$$
$$= \delta_{rs}\mathscr{E}(n_r) + \sum_{i,j,k}(-1)^k\frac{(N-r-i-k)!\,(N-s-j-k)!\,(r+i+k)!\,(s+j+k)!}{(N-r-s-k)!\,N!\,i!\,j!\,k!\,(s-i)!\,(r-j)!}$$
$$\times\mathscr{E}(n_{r+i+k})\mathscr{E}(n_{s+j+k}) - \sum_i(-1)^i\frac{(r+s+i)!}{r!\,s!\,i!}\mathscr{E}(n_{r+s+i}) \tag{28}$$
$$\simeq \delta_{rs}\mathscr{E}(n_r) + \sum_{i,j,k}(-1)^k\frac{(N-r-i-k)!\,(N-s-j-k)!\,(r+i+k)!\,(s+j+k)!}{(N-r-s-k)!\,N!\,i!\,j!\,k!\,(s-i)!\,(r-j)!}$$
$$\times\mathscr{E}(n_{r+i+k})\mathscr{E}(n_{s+j+k}) - 2^{-r-s}\frac{(r+s)!}{r!\,s!}\mathscr{E}(n_{r+s}(2)), \tag{29}$$

using (8). Provided i, j, k, r, s are all $\ll N$, the coefficient in the first sum is $O((rs/N)^{i+j+k})$; and when $i=j=k=0$, use of Stirling's formula shows that it is $1+O(rs/N)$. Hence, if $rs \ll N$,

$$\operatorname{cov}(n_r, n_s) = \mathscr{E}(n_r n_s) - \mathscr{E}(n_r)\mathscr{E}(n_s)$$
$$\simeq \delta_{rs}\mathscr{E}(n_r) - 2^{-r-s}\binom{r+s}{r}\mathscr{E}(n_{r+s}(2)). \tag{30}$$

Notice that when $r=s$ we have equation (22) of G:

$$V(n_r) \simeq \mathscr{E}(n_r) - 2^{-2r}\binom{2r}{r}\mathscr{E}(n_{2r}(2)), \tag{31}$$

or, expanding the second term by (8),

$$V(n_r) \simeq \mathscr{E}(n_r) - \sum_i(-1)^i\frac{(2r+i)!}{r!\,r!\,i!}\mathscr{E}(n_{2r+i}). \tag{32}$$

For the case $r=0$, since s is constant, we have

$$V(d) = V(n_0) \simeq \mathscr{E}(d(2)) - \mathscr{E}(d) \simeq \mathscr{E}(n_1) - \mathscr{E}(n_2) + \ldots. \tag{33}$$

Using (30), we can now find the variance of $\hat{n}_r(\lambda)$. From the elementary formula for the variance of a linear form: $V(\sum_i a_i x_i) = \sum_{i,j} a_i a_j \operatorname{cov}(x_i, x_j)$,

we have

$$
\begin{aligned}
&V(\hat{n}_r(\lambda)) \\
&\quad = \lambda^{2r} \sum_{i,j=0}^{\infty} (-1)^{i+j} (\lambda-1)^{i+j} \binom{r+i}{r} \binom{r+j}{r} \operatorname{cov}(n_{r+i}, n_{r+j}) \\
&\quad \doteq \lambda^{2r} \sum_{i,j=0}^{\infty} (-1)^{i+j} (\lambda-1)^{i+j} \binom{r+i}{r} \binom{r+j}{r} \left[\delta_{ij} \mathscr{E}(n_{r+i}) - 2^{-2r-i-j} \binom{2r+i+j}{r+i} \mathscr{E}(n_{2r+i+j}(2)) \right] \\
&\quad = \lambda^{2r} \left[\sum_{i=0}^{\infty} (\lambda-1)^{2i} \binom{r+i}{r}^2 \mathscr{E}(n_{r+i}) \right. \\
&\qquad \left. - \sum_{l=0}^{\infty} (-1)^l (\lambda-1)^l 2^{-2r-l} \mathscr{E}(n_{2r+l}(2)) \frac{(2r+l)!}{l!\,r!\,r!} \sum_{i,j \geqslant 0}^{i+j=l} \frac{(i+j)!}{i!\,j!} \right] \\
&\quad = \lambda^{2r} \left[\sum_{i=0}^{\infty} (\lambda-1)^{2i} \binom{r+i}{r}^2 \mathscr{E}(n_{r+i}) - \sum_{l=0}^{\infty} (-1)^l (\lambda-1)^l 2^{-2r} \frac{(2r+l)!}{r!\,r!\,l!} \mathscr{E}(n_{2r+l}(2)) \right] \\
&\quad \doteq \lambda^{2r} \left[\sum_{i=0}^{\infty} (\lambda-1)^{2i} \binom{r+i}{r}^2 \mathscr{E}(n_{r+i}) - \binom{2r}{r} (2\lambda)^{-2r} \mathscr{E}(n_{2r}(2\lambda)) \right], \qquad (34)
\end{aligned}
$$

using (8) again. This derivation of (34) is slightly unsatisfactory, since the second series in the previous formula may give a good approximation to the term by which we replace it only after so many terms have been taken that the approximation made in (30) is no longer valid. It may be possible to postpone making this approximation until a later stage of the calculation, but the algebra would become very heavy. In order to estimate $\mathscr{E}(n_{2r}(2\lambda))$ in calculating (34) for an actual case, it may be possible to use the method of § 2 (probably in conjunction with the summation technique described in § 5), or it may be easier to make a sufficiently accurate guess.

If $\hat{n}_r(\lambda)$ is defined by (18) with n'_{r+i} in place of n_{r+i}, it becomes very difficult to make any estimate of its variance. We can say, however, that so long as we feel that it is worth using smoothed values at all, the variance of the estimate based on them is likely to be considerably less than that given by (34).

4. Variance of $\hat{n}_r(\lambda)$ considered as a prediction of $n_r(\lambda)$

In the last section, we were considering the question: How much may $\hat{n}_r(\lambda)$ be expected to differ from its mean value (which is equal to $\mathscr{E}(n_r(\lambda))$)? A question which may sometimes be more relevant is: How much may $\hat{n}_r(\lambda)$ be expected to differ from the value of $n_r(\lambda)$ obtained in a random second sample? To answer this question, we want to find

$$V(\hat{n}_r(\lambda) - n_r(\lambda)), \qquad (35)$$

which may be called the *variance of $\hat{n}_r(\lambda)$ considered as a prediction of the random variable* $n_r(\lambda)$, rather than as an estimate of the parameter $\mathscr{E}(n_r(\lambda))$. It is evidently now necessary to consider separately the cases when the second sample is independent, and when it is an enlargement of the basic sample.

When the second sample is independent, $\hat{n}_r(\lambda)$ and $n_r(\lambda)$ are independent random variables, and so

$$V(\hat{n}_r(\lambda) - n_r(\lambda)) = V(\hat{n}_r(\lambda)) + V(n_r(\lambda)) \qquad (36)$$

which can be calculated by using (34) and the following modification of (31):

$$V(n_r(\lambda)) \doteq \mathscr{E}(n_r(\lambda)) - 2^{-2r} \binom{2r}{r} \mathscr{E}(n_{2r}(2\lambda)). \qquad (37)$$

In the case when the second sample is an enlargement, $\hat{n}_r(\lambda)$ and $n_r(\lambda)$ are correlated, and we have not been able to calculate (35) in this case. It may be expected to be considerably smaller than in the case of independence, at least if λ is not large, since when $\lambda = 1$ we have $\hat{n}_r(1) = n_r = n_r(1)$ and so (35) is reduced to zero.

5. Summation of the series obtained in §§2 and 3

We consider the case of the general series (18); similar remarks apply to (22), (23), (24) and (32), and to the series arising in calculating (34) and (37). It was pointed out in §1 that the term $(\lambda-1)^i$ in (8) and (18) may cause trouble if $\lambda > 2$. In fact, the series is likely to become 'practically divergent'; i.e. to behave like an infinitely oscillating series up to a point at which we become too uncertain of the value of $\mathscr{E}(n_r)$ to continue with the calculation. This difficulty is illustrated by formula (34) for the variance of $\hat{n}_r(\lambda)$; if $\lambda > 2$, the series

$$\sum_{i=0}^{\infty} \binom{r+i}{r}^2 (\lambda-1)^{2i}\, \mathscr{E}(n_{r+i})$$

is likely to have a very large sum, unless $\mathscr{E}(n_{r+i})$ decreases extremely fast. It is natural to try to overcome this difficulty by using a method of summation which is known to make some oscillating series converge; a convenient method appears to be that of Euler, with a parameter q, generally called the (E,q) method (Hardy, 1949, pp. 178 ff.). This is to transform the series $\sum_{i=0}^{\infty} a_i$ into $\sum_{j=0}^{\infty} a_j^{(q)}$, where

$$a_j^{(q)} = \frac{1}{(q+1)^{j+1}} \sum_{i=0}^{j} \binom{j}{i} q^{j-i} a_i \tag{38}$$

$$= (-1)^j \left(\frac{q}{q+1}\right)^{j+1} \Delta_0^j \left[(-1)^i \frac{a_i}{q^{i+1}}\right], \tag{39}$$

the forward difference symbol Δ_r^j being defined inductively by

$$\Delta_r^1 \alpha_i = \alpha_{r+1} - \alpha_r, \quad \Delta_r^j = \Delta_r^1 \Delta_i^{j-1}. \tag{40}$$

(The form (39) is given by Bromwich (1926), pp. 62–6, for the case $q = 1$. It leads to a convenient method of setting out the work in a practical example, which will be illustrated in Example (i) of the next section.) If $\sum_i a_i$ converges, then $\sum_j a_j^{(q)}$ converges to the same sum, for any $q \geqslant 0$; if $\sum_j a_j^{(q')}$ converges, then $\sum_j a_j^{(q)}$ converges to the same sum for all $q \geqslant q'$ (Hardy, 1949, Theorems 117 and 118).

In practical examples, n_{r+i} generally decreases slowly after the first few terms, and we are usually interested in small values of r, so that $\binom{r+i}{r}$ increases slowly. Under these circumstances, the series (18) is, after the first few terms, nearly a G.P. with ratio $-(\lambda-1)$. Now, if we apply the (E,q) method to such a G.P., say

$$a_i = (-1)^i (\lambda-1)^i a_0, \tag{41}$$

we obtain

$$a_j^{(q)} = \frac{a_0}{(q+1)^{j+1}} \sum_{i=0}^{j} \binom{j}{i} q^{j-i} (-(\lambda-1))^i$$

$$= \frac{a_0}{(q+1)^{j+1}} (q-(\lambda-1))^j$$

$$= \frac{a_0}{q+1} \left(\frac{q-(\lambda-1)}{q+1}\right)^j, \tag{42}$$

i.e. the transformed series is a G.P. with ratio $\frac{q-(\lambda-1)}{q+1}$. Clearly the best value of q to select is $\lambda-1$, which reduces all but the first term of the transformed series to zero.† If the n_{r+i} decrease fairly rapidly, we may get better results by choosing q somewhat smaller. (This is the case in Example (i) of §6, where $\lambda = 5$, but we take $q = 2$.) When $r = 0$ or 1, and if $n_1 \gg n_2$, it may be worth taking out the first term of the series, and applying the summation process to the remainder; the reason for this can be seen by considering such a series as

$$1-\tfrac{1}{8}+\tfrac{1}{4}-\tfrac{1}{2}+\ldots. \tag{43}$$

If we apply the $(E, 2)$ method directly, we get

$$0{\cdot}333+0{\cdot}208+0{\cdot}139+0{\cdot}093+\ldots, \tag{44}$$

while if we take out the first term and apply the $(E, 2)$ method to the remainder, we get

$$1{\cdot}000-0{\cdot}042+0{\cdot}000+0{\cdot}000+\ldots, \tag{45}$$

which is evidently better.

When we have chosen a method of summation, and selected a partial sum of the transformed series as probably giving a sufficiently good approximation to the final sum, we can express this partial sum as a linear combination of the n_r, and deduce its variance, as in §3.‡ But there is now a new source of error, namely, the omission of the rest of the transformed series. We have not been able to find a useful form of error term for this remainder (corresponding to the statement that alternate partial sums of (8) err in excess and defect); failing such an error term, our results must be used with caution when it is necessary to apply the summation process. If the n_r's decrease slowly and q is taken to be slightly smaller than $\lambda-1$, the transformed series will generally have terms alternating in sign (cf. equation (42)); it might then be hoped that the partial sums err alternately in excess and defect, but it does not seem to be possible to lay down any simple general conditions under which this is the case.

Some of the methods described by Shanks (1955) also seem to be very well suited to our case; they have the property of summing perfectly any series which is geometric from some point onwards, so that the difficulty caused by an excessively large first term, noted above, does not arise. Given the series $\sum_{n=0}^{\infty} a_n$ we define a sequence (not a new series) by

$$B_n = \sum_{r=0}^{n-1} a_r + \frac{a_n^2}{a_{n+1}-a_n} \quad (n = 1, 2, 3, \ldots);$$

repetition of the process gives a sequence C_n, and so on. The e_1 method consists of considering the sequence B_n (in place of the sequence of partial sums $A_n = \sum_{r=0}^{n} a_r$), the e_1^2 method, of considering the sequence C_n, and so on; the $\tilde{e}_1$ method consists of considering the sequence $A_0, B_1, C_2, \ldots$. For an example, see §6, Example (i).

† This statement may appear to conflict with the remark of Hardy (1949), p. 180, that 'as q increases, the (E, q) methods form a scale of increasing strength'. But here 'strength' refers only to whether we obtain a convergent series or not: if we choose q unnecessarily large, we shall certainly obtain a convergent series, but it will converge very slowly.

‡ This remark applies to any method of summation by a linear transformation of the series, e.g. Cesàro means, the composite $(E, q; C, k)$ method, any Hausdorff means, Hölder means, Hutton's method, any Nörlund means, or quasi-Hausdorff transformations; for references to all these methods, see Hardy (1949), p. 392. It does not apply to the non-linear methods of Shanks (1955), mentioned below.

6. EXAMPLES

The first example is an artificial one designed to test the efficacy of the methods described above, especially the summation methods of §5. The second and third examples illustrate the practical applications, but enlarged samples are not available for verifying the estimates.

Example (*i*). *Sample of words from 'Our Mutual Friend' by Charles Dickens.* The following samples were taken:

A, of 1000 words, the last words of lines on pages $\equiv 5 \bmod 25$,

B, of 2000 words, the last words of lines on pages $\equiv 10$ or $20 \bmod 25$,

C, of 2000 words, the last words of lines on pages $\equiv 15$ or $25 \bmod 25$,

the sampling in each case being carried as far as required to make up the prescribed number of words. Our original intention was to use A as the basic sample ($N = 1000$) and to calculate

Table 1

Sample A; $N = 1000$			Sample A; $N = 1000$		
r	n_r	n'_r	r	n_r	n'_r
1	404	404	6	3	—
2	57	64	7	0	—
3	24	25	8	3	—
4	16	12·2	$\geqslant 9$	15	—
5	6	6·2			
$d = 528$					

values of $\hat{d}(\lambda)$ and $\hat{n}_1(\lambda)$ given by (19) and (18) for $\lambda = 2, 3, 4, 5$, which could be checked against the values of $d(\lambda)$ and $n_1(\lambda)$ actually obtained from the samples B, $A+B$, $B+C$, $A+B+C$. The results, however, showed a systematic and, for $\hat{d}(2)$, significant difference between the prediction and the observed result. Working back from sample B with $\lambda = \frac{1}{2}$, it appeared that sample A had n_1 considerably too small. We believe that this is due to the fact that the method of sampling used was not sufficiently random; an uncommon word is likely to occur several times on the same page, where a particular topic is discussed, and such a word is therefore less likely to occur just once in a sample selected as described than in a random sample of the same size.†

The results for larger values of λ were, however, not much less accurate than those for $\lambda = 2$, and we give the calculation of $\hat{d}(5)$ as an example of the use of the (E, q) method of summation described in §5. Table 1 shows the data; the n'_r were obtained by graphical smoothing of $\sqrt{n_r}$. Our formula (19) gives us

$$\hat{d}(5) = 528 + (4.404 - 4^2.64 + 4^3.25 - 4^4.12{\cdot}2 + 4^5.6{\cdot}2 - \ldots). \tag{46}$$

† Consider the extreme case when p. 1 reads 'one one one...', p. 2 'Two two two...', and so on.

To transform the bracketed series we form the difference table suggested by (39) (q has been chosen as 2, so as to make the differences small):

$$\begin{array}{llllll}
\frac{1}{2}.4.404 & =808 & & & & \\
 & & -552 & & & \\
(\frac{1}{2})^2.4^2.64 & =256 & & 496 & & \\
 & & -56 & & -445 & \\
(\frac{1}{2})^3.4^3.25 & =200 & & 51 & & 402 \\
 & & -5 & & -43 & \\
(\frac{1}{2})^4.4^4.12{\cdot}2 \doteq & 195 & & 8 & & \\
 & & 3 & & -\overline{488} & \\
(\frac{1}{2})^5.4^5.6{\cdot}2 \doteq & 198 & & \overline{555} & & \\
 & & -\overline{610} & & &
\end{array}$$

(We apply the usual check, that the sum of each column is equal to the difference between the top and bottom of the one before.) The transformed series, by (39), is

$$\tfrac{2}{3}.808+(\tfrac{2}{3})^2.552+(\tfrac{2}{3})^3.496+(\tfrac{2}{3})^4.445+(\tfrac{2}{3})^5.402\ldots \doteq 538+245+147+88+53\ldots. \quad (47)$$

The last few terms of (47) are approximately a geometric series with ratio 0·6; the sum of the remaining terms should therefore be approximately†

$$53\times\frac{0{\cdot}6}{1-0{\cdot}6}\doteq 79,$$

making a total of 1150; hence

$$\hat{d}(5) = 528+1150 = 1678. \quad (48)$$

Applying the methods of Shanks (1955), described at the end of §5, we get Table 2. Although the transformed sequences are rather short, it looks as if $C_2 = 1155$ is a good approximation to the limit, giving $\hat{d}(5) = 1683$. In fact, for the whole sample $A+B+C$, $d(5) = 1832$.

Table 2

n	A_n	B_n	C_n
0	1616		
1	592	1216	
2	2192	1134	1155
3	−931	1162	
4	5418		

In this example, we have been slightly handicapped by having so few terms of the series available; when using the (E, k) method, this renders the remainder somewhat uncertain, and prevents us from omitting the first term from the summation process, as was suggested in §5. Because of this difficulty and the apparent non-randomness noted above, we use sample B ($N = 2000$) as the basic sample for a more comprehensive test of our methods, although we can then verify the results only up to $\lambda = 2{\cdot}5$. (The 'second samples' for $\lambda = 1{\cdot}5$, 2·0, 2·5 are $A+B$, $B+C$, $A+B+C$ respectively, and are thus all enlargements of the basic sample.) Table 3 gives the data for this basic sample; the n'_r were produced

† This is, as Shanks (1955) points out, equivalent to applying the e_1 method to sum (47).

by smoothing $\sqrt{n_r}$ graphically by the use of French curves, the n_r'', independently, by smoothing by eye: $\hat{d}(\lambda)$, $\hat{n}_1(\lambda)$, and the estimated percentage coverage, $100\,(1-\hat{n}_1(\lambda)/\lambda N)$, were calculated for $\lambda = 1{\cdot}5,\ 2{\cdot}0,\ 2{\cdot}5$, using the three sets of values. The summation process was used only in the case $\lambda = 2{\cdot}5$, with $q = 1$. Table 4 shows the three sets of estimates and the actual results found in the enlarged samples; standard deviations are given where applicable, calculated from (31), (33), and (34). It will be noticed that in this case little or nothing was gained when smoothed values were used; but it would probably be essential to use smoothed values when working with larger values of λ.

Table 3

Sample B; $N = 2000$			
r	n_r	n_r'	n_r''
1	729	729	729
2	108	96	110
3	33	38	38
4	23	21	19
5	17	14	13
6	7	9	9
7	5	7	6
8	3	3·2	4
$\geqslant 9$	30	—	—
$d = 955$			

Example (*ii*). *Captures of Macrolepidoptera in a light-trap at Rothamsted.* (Quoted as example (i) in G, § 8 from Williams's data in Corbet, Fisher & Williams (1943).) $N = 15609$, $d = 240$. Table 5 shows the small values of r. n_r' is n_r''' of the example in G, obtained by smoothing $\sum_{t=1}^{r} tn_t$. n_r'' is Fisher's analytic smoothing, given by H_3 of G with parameter $\beta = 40{\cdot}2$. Now H_3 is a hypothesis defining the distribution of the population frequencies $\{p_\mu\}$, and it implies that

$$\mathscr{E}(n_r(\lambda)) \simeq \frac{\beta}{r}\left(\frac{N\lambda}{N\lambda+\beta}\right)^r. \tag{49}$$

(G, (63)), and

$$\mathscr{E}(d(\lambda)) = \beta \log_e\left(\frac{N\lambda}{\beta}+1\right) \tag{50}$$

(G, (67)). Since $N \gg \beta$, we see that H_3 implies

$$\mathscr{E}(n_1(\lambda)) \simeq \mathscr{E}(n_1) \tag{51}$$

and

$$\mathscr{E}(d(\lambda)) \simeq d + \beta \log_e \lambda. \tag{52}$$

Putting $\lambda = 2$ we see that doubling the sample will approximately halve the proportion of the population not represented (by (51) and (3)) and increase the number of distinct species observed by approximately $\beta \log_e 2 = 27{\cdot}9$. (The latter fact was noted by Williams in Corbet *et al.* (1943), p. 51.)

Table 4

Estimates of $d(\lambda)$	$\lambda = 1{\cdot}5$	$\lambda = 2{\cdot}0$	$\lambda = 2{\cdot}5$
Using n_r	1296 ± 34	1599 ± 50	1872
Using n'_r	1299	1613	1909
Using n''_r	1296	1601	1890
Actual results	1303 ± 29	1551 ± 31	1832 ± 34

Estimates of $n_1(\lambda)$	$\lambda = 1{\cdot}5$	$\lambda = 2{\cdot}0$	$\lambda = 2{\cdot}5$
Using n_r	958 ± 42	1172†	1322
Using n'_r	981	1238	1435
Using n''_r	961	1168	1350
Actual results	983 ± 29	1116 ± 31	1308 ± 32

Estimates of % coverage	$\lambda = 1{\cdot}5$	$\lambda = 2{\cdot}0$	$\lambda = 2{\cdot}5$
Using n_r	$68{\cdot}1 \pm 1{\cdot}4$	70·7†	73·6
Using n'_r	67·3	69·0	71·3
Using n''_r	68·0	70·8	73·0
Actual results	$67{\cdot}2 \pm 1{\cdot}0$	$72{\cdot}1 \pm 0{\cdot}8$	$73{\cdot}8 \pm 0{\cdot}6$

† The S.D. is not given in this case because the sum of the series was not taken to infinity but estimated after eight terms as lying midway between the last partial sums; that is, in effect, Hutton's method (Hu, 1) was applied to sum the series (Hardy, 1949, pp. 21–2). (34) would give a very large variance, most of which arises from terms after the eighth.

Table 5

r	n_r	n'_r	n''_r
1	35	35·0	40·1
2	11	22·5	20·0
3	15	16·3	13·3
4	14	12·3	10·0
5	10	9·7	7·9
6	11	7·7	6·6
7	5	6·0	5·6

The present example is not a good one to which to apply our distribution-free† methods, since even n_1 is rather small; however, we shall obtain the corresponding results for comparison, using the smoothed values n'_r. By (18),

$$\begin{aligned}\hat{n}_1(2) &\simeq 2(35{\cdot}0 - 45{\cdot}0 + 48{\cdot}9 - 49{\cdot}2 + 48{\cdot}5 - 46{\cdot}2 + 42{\cdot}0 - \ldots) \\ &= 2(17{\cdot}5 - 2{\cdot}5 - 0{\cdot}8 - 0{\cdot}2 + 0{\cdot}0 + 0{\cdot}1 + 0{\cdot}1 + \ldots)\end{aligned}$$

(transforming the series by $(E, 1)$)

$$= 28{\cdot}4,$$

whence (using (3)) we predict that the proportion of the population not covered should decrease from

$$\frac{35{\cdot}0}{15609} = 0{\cdot}22\ \%$$

to

$$\frac{28{\cdot}4}{31218} = 0{\cdot}09\ \%,$$

whereas H_3 implied that it was halved. By (24), the expected number of new species is approximately

$$\begin{aligned}&35{\cdot}0 - 22{\cdot}5 + 16{\cdot}3 - 12{\cdot}3 + 9{\cdot}7 - 7{\cdot}7 + 6{\cdot}0 - \ldots \\ &\quad = 17{\cdot}5 + 3{\cdot}1 + 0{\cdot}8 + 0{\cdot}3 + 0{\cdot}1 + 0{\cdot}1 + 0{\cdot}0 + \ldots \quad \text{(by } (E, 1)) \\ &\quad = 21{\cdot}9,\end{aligned}$$

whereas H_3 implied that 27·9 were expected. Finally, in order to estimate the S.D. of n_1 we calculate from (18)

$$\begin{aligned}\hat{n}_2(2) &\simeq 4(22{\cdot}5 - 48{\cdot}9 + 73{\cdot}8 - 97{\cdot}0 + 115{\cdot}5 - 126{\cdot}0 + \ldots) \\ &= 4(11{\cdot}25 - 6{\cdot}60 - 0{\cdot}19 + 0{\cdot}01 - 0{\cdot}09 - 0{\cdot}04 + \ldots) \quad \text{(by } (E, 1)) \\ &= 17{\cdot}4,\end{aligned}$$

so, by (31),

$$V(n_1) \simeq 35 - \tfrac{1}{2}.2.17{\cdot}4 = 26{\cdot}3,$$

giving n_1 a S.D. of about 5·1, whereas H_3 (by (65) of G) gave 5·5.

We note that our distribution-free estimates of $n_1(2)$, $n_2(2)$, $d(2)$ are all less than the values implied by H_3. This is what one would expect when the sample size, N, is large enough for the finiteness of s (the total number of species in the population) to conflict with the prediction of H_3 that $s = \infty$; but the effect may well be accidental.

In this example, we can also make a rough comparison between the variances of the estimate of $n_1(\lambda)$ deduced from H_3 and that given by our distribution-free method. The former estimate is almost exactly β, and its variance is therefore approximately equal to the sampling variance of β for this example, given by Fisher (Corbet *et al.* (1943), p. 56) as 1·13 (S.D. of 1·06); this is independent of λ. Our $\hat{n}_1(\lambda)$, on the other hand, has a very large variance for $\lambda \geqslant 2$ (by (34)), and even for $\lambda = 1$, since $\hat{n}_1(1) = n_1$, its variance, as we saw above, is about 26·3 (S.D. 5·1). However, it must be remembered, first, that H_3 is certainly not exactly true (since in fact $s < \infty$),‡ so that the estimate deduced from it is subject to an unknown additional error, and secondly, that we may hope to reduce the variance of $\hat{n}_r(\lambda)$ considerably by using carefully smoothed values and summation methods.

† I.e. independent of any particular assumption about the distribution of the p_μ, in contrast to the above argument which assumes H_3.

‡ Not even the truncated form, H_5 of G, can be exactly true, as was shown in G, p. 257.

In general, if a fairly simple hypothesis H on the p_μ (e.g. any of H_1 to H_6 of G) gives a good fit for the n_r, we should prefer to deduce $\mathscr{E}(n_r(\lambda))$ and $\mathscr{E}(d(\lambda))$ from H, rather than use the distribution-free estimates (18) and (19); but such extrapolation should be made with caution, and the distribution-free methods may give a useful indication of the error to be expected if H is false.

Example (*iii*). *Sample of nouns in Macaulay's essay on Bacon.* (From Yule (1944), Table 44, p. 163; quoted in G as example (iii), p. 260.) $N = 8045$, $d = 2048$.

Table 6

r	n_r	n'_r	r	n_r	n'_r
1	990	1024	11	24	15·5
2	367	341	12	19	13·1
3	173	170	13	10	11·3
4	112	102	14	10	9·7
5	72	68	15	13	8·5
6	47	49	16–20	31	30·5
7	41	35·5	21–30	31	31·5
8	31	28·5	31–50	19	25·9
9	34	22·7	51–100	6	19·9
10	17	18·4	101–∞	1	20·3

(As in the tables in G, n_r and n'_r have been summed where values of r are grouped.)

Here n'_r ($= n''_r$ of G) is the analytic smoothing

$$n'_r = \frac{2048}{r(r+1)}. \tag{53}$$

(H_9 of G; notice that this is *not* an explicit hypothesis on the p_μ, and that it gives a good fit only for $r \leqslant 30$.) (53) is so simple in form that we can carry through all our calculations analytically. Again we consider doubling the sample ($\lambda = 2$); by (18),

$$\begin{aligned}\hat{n}_1(2) &= 2\sum_{i=0}^{\infty}(-1)^i(i+1)\frac{2048}{(i+1)(i+2)}\\ &= 2.2048\sum_{i=0}^{\infty}\frac{(-1)^i}{i+2}\\ &= 2.2048(1-\log_e 2)\\ &\simeq 1260;\end{aligned}$$

and

$$\begin{aligned}\hat{n}_2(2) &= 4\sum_{i=0}^{\infty}(-1)^i\frac{(i+1)(i+2)}{2}\frac{2048}{(i+2)(i+3)}\\ &= 2.2048\sum_{i=0}^{\infty}(-1)^i\frac{i+1}{i+3}\\ &= 2.2048\left[\sum_{i=0}^{\infty}(-1)^i - 2\sum_{i=0}^{\infty}\frac{(-1)^i}{i+3}\right].\end{aligned}$$

The first series is summed by any standard method to $\frac{1}{2}$, and the second is equal to $\log_e 2 - \frac{1}{2}$; hence

$$\begin{aligned}\hat{n}_2(2) &= 2.2048(\tfrac{3}{2} - 2\log_e 2)\\ &\simeq 465.\end{aligned}$$

Finally, by (19),

$$\begin{aligned}\hat{d}(2) &= 2048 - \sum_{i=1}^{\infty} (-1)^i \frac{2048}{i(i+1)} \\ &= 2048\left[1 - \sum_{i=1}^{\infty}\left(\frac{(-1)^i}{i} + \frac{(-1)^{i+1}}{i+1}\right)\right] \\ &= 2048.2\sum_{i=1}^{\infty}\frac{(-1)^{i+1}}{i} \\ &= 2048.2\log 2 \\ &\simeq 2840.\end{aligned}$$

Notice that, since the n'_r give a good fit only for $r \leqslant 30$, the justification for substituting them in infinite series rests on the following argument:

(i) the partial sum of the series† with the true $\mathscr{E}(n_r)$ down to the term containing $\mathscr{E}(n_{30})$ is a good approximation to its infinite sum;

(ii) the same is true for the series† with the n'_r;

(iii) the n'_r are good approximations to the $\mathscr{E}(n_r)$ for $r \leqslant 30$, so that the partial sums mentioned are nearly equal.

(Compare the argument justifying evaluation of integrals by the saddle-point method.) (i)–(iii) should be borne in mind whenever an analytic smoothing is used in this way, or even when it is used to give values of n'_r which are treated numerically; otherwise there is some risk of obtaining an apparently satisfactory convergence which is in fact spurious. When possible, it would probably be advisable in such cases to try a graphical smoothing as well: the reader might like to try the smoothing n'_r of G, using the $(E, 1)$ or $\tilde{e}_1$ method of summation.

We have now sufficient data to derive the result which was quoted in §1. By (3), the proportion of the population not represented in the 2048 nouns of the basic sample is about‡

$$\tfrac{1024}{8045} \simeq 12{\cdot}7\ \%;$$

by (7) of G the proportion not represented in the 2840–1260 = 1580 nouns occurring twice or more in the doubled sample will be about

$$\frac{1260 + 2.465}{16090} \simeq 13{\cdot}6\ \%;$$

by (2) of G, the average frequency of the 1260 nouns occurring once only in the doubled sample will be about

$$\frac{2.465}{1260.16090} \simeq 0{\cdot}0046\ \%.$$

Hence, if we add a random selection of

$$\frac{13{\cdot}6 - 12{\cdot}7}{0{\cdot}0046} \simeq 200$$

of the nouns occurring once only in the doubled sample to all those occurring twice or more, we will have a list of about 1780 nouns covering approximately the same proportion of the population as the 2048 nouns of the basic sample.

† Or, if summation methods are used (as in calculating $\hat{n}_2(2)$ above), the sum of the transformed series.

‡ The figure of 12·3 % given in G was based on the unsmoothed values.

APPENDIX

Conditions for the lemma of §1

Although this lemma was not actually used in our argument, we give here, for any reader who may be interested, two fairly general sets of conditions under which it holds.

If $a_r \geqslant 0$ for all r, and finite numbers b_i are defined by (4), *then* (5) *holds if and only if*

$$\frac{b_{r+i}}{i!} \to 0 \quad \text{as} \quad i \to \infty. \tag{54}$$

Proof. Write

$$R(n,r) = \frac{1}{r!}\sum_{i=0}^{n}\frac{(-1)^i b_{r+i}}{i!} - a_r,$$

so that (5) holds if and only if $R(n,r) \to 0$ as $n \to \infty$. Now, for all $n > r$,

$$\begin{aligned} R(n,r) &= \frac{1}{r!}\sum_{i=0}^{n}\frac{(-1)^i}{i!}\sum_{s=0}^{\infty} s^{(r+i)}a_s - a_r \\ &= \sum_{s=0}^{\infty}\binom{s}{r}a_s\sum_{i=0}^{n}(-1)^i\binom{s-r}{i} - a_r \\ &= \sum_{s=0}^{\infty}(-1)^n\binom{s}{r}\binom{s-r-1}{n}a_s - a_r, \end{aligned}$$

using the definition $\binom{a}{b} = \frac{a^{(b)}}{b!}$, even if a is negative, together with the well-known identity

$$\binom{s-r}{i} = \binom{s-r-1}{i-1} + \binom{s-r-1}{i}.$$

Putting $s = r$ gives a term $+a_r$, and all other terms with $s < n+r+1$ vanish; hence

$$\begin{aligned} R(n,r) &= (-1)^n\sum_{s=n+r+1}^{\infty}\binom{s}{r}\binom{s-r-1}{n}a_s \\ &= (-1)^n\sum_{s=n+r+1}^{\infty}\frac{s-r-n}{s-r}\frac{s^{(n+r)}}{n!\,r!}a_s. \end{aligned} \tag{55}$$

Now, since $a_s \geqslant 0$ for all s,

$$\begin{aligned} |R(n,r)| &\leqslant \frac{1}{n!\,r!}\sum_{s=0}^{\infty}s^{(n+r)}a_s \\ &= \frac{1}{r!}\frac{b_{r+n}}{n!}; \end{aligned}$$

and the sufficiency of (54) follows. The necessity is trivial, since if (54) does not hold the right-hand side of (5) cannot converge.

If $a_r = O(x^r)$, $0 \leqslant x < \frac{1}{2}$, then (5) *holds for all r; further,* (5) *does not hold if $a_r = 2^{-r}$, so this result cannot be improved by extending the range of x.*

Proof. (i) We may assume without loss of generality that $|a_r| \leqslant x^r$. Then it follows from (55) above that

$$\begin{aligned} |R(n,r)| &\leqslant \sum_{s=n+r}^{\infty}\frac{s^{(n+r)}}{n!\,r!}x^s \\ &= \frac{(n+r)!\,x^{n+r}}{n!\,r!}\sum_{t=0}^{\infty}\binom{-n-r-1}{t}(-x)^t \\ &= \frac{1}{r!\,x}(n+r)^{(r)}\left(\frac{x}{1-x}\right)^{n+r+1} \\ &\to 0 \text{ as } n \to \infty \text{ provided } x < \tfrac{1}{2}; \end{aligned}$$

hence (5) holds for all r.

(ii) Taking $a_r = 2^{-r}$, we have

$$\begin{aligned} b_i &= \sum_{r=0}^{\infty}r^{(i)}(\tfrac{1}{2})^r \\ &= 2.i!, \end{aligned}$$

summing the series as in (i), and it is clear that the right-hand side of (5) is not convergent for any r.

REFERENCES

BROMWICH, T. J. I'A. (1926). *An Introduction to the Theory of Infinite Series*, 2nd ed. Cambridge University Press.

CORBET, A. S., FISHER, R. A. & WILLIAMS, C. B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* **12**, 42–58.

DICKENS, CHARLES. *Our Mutual Friend.* London: Thomas Nelson. (First published in 1864–5.)

GOOD, I. J. (1953) (described in text as G). The population frequencies of species and the estimation of population parameters. *Biometrika*, **40**, 237–64.

HARDY, G. H. (1949). *Divergent Series.* Oxford: Clarendon Press.

SHANKS, D. (1955). Nonlinear transformations of divergent and slowly convergent series. *J. Math. Phys.* **34**, 1–42.

YULE, G. U. (1944). *Statistical Study of Literary Vocabulary.* Cambridge University Press.

A SEQUENTIAL TEST OF RANDOMNESS FOR EVENTS OCCURRING IN TIME OR SPACE

By D. J. BARTHOLOMEW

University College London and Scientific Department, National Coal Board, London

1. Introduction

In a variety of practical problems it is necessary to test whether a sequence of events is occurring at random in time or space. To the casual observer even a random series will suggest fluctuations in the density of events, so that an objective test is clearly required to establish randomness or otherwise.

The problem was discussed by Maguire, Pearson & Wynn (1952) as applied to industrial accidents; they pointed out that the basic data in such cases consisted of an ordered sequence of intervals between events. Recent investigations into the possible departures from randomness have shown that it is sometimes appropriate to use a test based on intervals, but on other occasions the times form the best starting point. In this paper we give a sequential test which utilizes the times at which events occur because the alternative envisaged requires this approach. A sequential test is of special value in this case because the observations become available one at a time; provided that they do not follow each other too rapidly it should be possible to carry out a test as the process develops.

2. The alternatives to randomness

In order to derive a sequential test we must first of all define the class of alternatives to which the test must be sensitive; this choice, of course, will be governed by the problem in view. Where the events are accidents or machine failures, for example, it will be most important to detect a systematic increase or decrease in the rate at which events are occurring; minor fluctuations are relatively unimportant. The test to be put forward is designed to be sensitive to the presence of this type of alternative.

We shall consider the class of processes for which

$$\Pr\{\text{event in } (T, T+dT)\} = \lambda(T)\,dT + o(dT) \quad (\lambda(T) \geqslant 0 \text{ for } T \geqslant 0),$$

where T is time (or distance) measured from the point at which observation was begun and $\lambda(T)$ is a monotonic function of T. The case, $\lambda(T) =$ constant, yields the familiar Poisson process; this will be the null hypothesis corresponding to randomness denoted by H_0. It will be convenient to refer to $\lambda(T)$ as the rate at which events are occurring at time T. The problem thus becomes that of testing the hypothesis, $\lambda(T) =$ constant, against the alternative, $\lambda(T)$ increasing or decreasing. As an example of a situation where this sort of alternative would be suitable, the data given by Maguire *et al.* (1952, Table 1), concerning mining accidents, may be mentioned. Accident and failure data, already referred to, furnish many examples of this type.

The raw material for the test consists of the times at which events are observed to occur, $T_1, T_2, \ldots, T_n$, and the first step must be to find their joint distribution. The distribution of

T_i clearly depends only on that of T_{i-1}, so that $p(T_1, T_2, \ldots, T_n)$ may be obtained from the relation

$$p(T_1, T_2, \ldots, T_n) = \prod_{i=1}^{n} p(T_i \mid T_{i-1}), \quad T_0 = 0.$$

Now
$$p(T_i \mid T_{i-1}) = \lambda(T_i) \exp\left[-\int_{T_{i-1}}^{T_i} \lambda(T)\, dT\right],$$

therefore

$$p(T_1, T_2, \ldots, T_n) = \prod_{i=1}^{n} \lambda(T_i) \exp\left[-\int_{0}^{T_n} \lambda(T)\, dT\right]. \tag{1}$$

In order that this can represent a joint-probability density function a further condition must be imposed on $\lambda(T)$, namely,

$$\lim_{x \to \infty} \int_0^x \lambda(T)\, dT = \infty.$$

This ensures that $\lambda(T)$ does not decay too rapidly; it is always satisfied if $\lambda(T)$ is monotonic increasing.

Before the Wald test can be derived $\lambda(T)$ must be completely specified; it will need to involve a scale factor as well as some parameter which determines the degree of departure from randomness. Without going farther than requiring that $\lambda(T)$ shall be monotonic decreasing or increasing a wide choice exists among the possible functional forms. It therefore seems reasonable to investigate the consequences of choosing that form which leads to the most straightforward mathematical treatment. For *mathematical* convenience therefore we choose

$$\lambda(T) = \mu(\mu T)^a \quad (a > -1).$$

This is monotonic decreasing for $-1 < a < 0$ and increasing for $0 < a < \infty$; the problem finally becomes that of testing the hypothesis $a = 0$ against either or both the alternatives $a > 0$ or $a < 0$. μ is a nuisance parameter which will be eliminated from the problem.

The important question which now arises concerns how far the test derived on this assumption will be efficient when the true alternative takes some other form. For example, it will be noted that the form proposed implies that when $T = 0$, $\lambda(T) = 0$ if $a > 0$ and $\lambda(T) = \infty$ if $a < 0$. Thus taken rigorously the origin chosen for T is of considerable importance. Although therefore the test based on the alternative $\lambda(T) = \mu(\mu T)^a$ will not always be the best possible, there are good reasons for believing that it will nevertheless be a good one under a variety of conditions. This matter will be discussed below in connexion with the practical application of the test. It may be noted that a detailed investigation carried out in fixed sample theory (Bartholomew, 1955) shows that the statistic $Q_1 = -2\Sigma \log T_i$ for this alternative form of $\lambda(T)$ has a fairly high efficiency under a wide range of other alternatives. It seems reasonable, therefore, to assume that the sequential version of the test will have this property.

For the joint density function of the T's under the chosen alternative, we find

$$p(T_1, T_2, \ldots, T_n) = \mu^n \prod_{i=1}^{n} (\mu T_i)^a \exp\left[-(\mu T_n)^{a+1}/(a+1)\right]. \tag{2}$$

It will often be desirable to have a test which can be applied when the first few observations are missing or have been used to reach earlier decisions. For this reason it will be more convenient to work with the more general density function of $T_k, T_{k+1}, \ldots, T_n$ $(k \geqslant 1)$ which

may be obtained by integrating out $T_1, \ldots, T_{k-1}$ over the region $0 \leqslant T_1 \leqslant T_2 \leqslant \ldots \leqslant T_{k-1} \leqslant T_k$. We find

$$p(T_k, T_{k+1}, \ldots, T_n) = \frac{\mu^{n-k+1}}{(k-1)!} \frac{(\mu T_k)^{(k-1)(a+1)}}{(1+a)^{k-1}} \prod_{i=k}^{n} (\mu T_i)^a \exp\left[-(\mu T_n)^{a+1}/(a+1)\right]. \tag{3}$$

3. The sequential Q-test

The parameter μ is of no interest and can be eliminated from the problem by carrying out a sequential transformation of the observations. Any transformation will lead to the same test provided it satisfies two conditions:

(i) It must be possible to form the new variables in a sequential manner.

(ii) The transformed variables must be independent of μ.

Probably the simplest transformation satisfying the required conditions is as follows:

$$v_k = \frac{T_k^k}{T_{k+1}^k}, \quad v_{k+1} = \frac{T_k^k T_{k+1}}{T_{k+2}^{k+1}}, \quad \ldots, \quad v_{n-1} = \frac{T_k^k T_{k+1} \ldots T_{n-1}}{T_n^{n-1}},$$

$$V = \mu T_n.$$

The test is based on the new variables $v_k, v_{k+1}, \ldots, v_{n-1}$. The Jacobian of the transformation is of the form $V^{n-1}/\mu^{n-k+1}\phi_1(v_k, \ldots, v_{n-1})$, where ϕ_1 is a function of the v's not depending on a. We thus obtain the joint density of $v_k, v_{k+1}, \ldots, v_{n-1}$ and V as

$$p(v_k, \ldots, v_{n-1}, V) = \frac{(v_{n-1})^a}{(k-1)!\,(1+a)^{k-1}} \phi_1(v_k, \ldots, v_{n-1}) \phi_2(v_k, \ldots, v_{n-1}) V^{n(a+1)-1} \exp -\left[V^{a+1}/(a+1)\right],$$

where $\phi_2 = (\mu T_k)^{k-1}$ can be expressed as a function of the v's which again is independent of a.

Integrating out for V from 0 to infinity, we finally obtain

$$p(v_k, \ldots, v_{n-1}) = (1+a)^{n-k}\, v_{n-1}^a \frac{(n-1)!}{(k-1)!} \phi_1 \phi_2. \tag{4}$$

It is now possible to form the sequential probability ratio test based on the $(n-k)$ variables $v_k, v_{k+1}, \ldots, v_{n-1}$. Thus

$$R = \frac{p(v_k, \ldots, v_{n-1} \,|\, a = a_0)}{p(v_k, \ldots, v_{n-1} \,|\, a = 0)} = (1+a_0)^{n-k}\, v_{n-1}^{a_0},$$

or taking logarithms and reverting to the original observations

$$\log R = (n-k)\log(1+a_0) - a_0 Q(n,k),$$

where

$$Q(n,k) = -\log v_{n-1} = (n-1)\log T_n - (k-1)\log T_k - \sum_{i=k}^{n-1} \log T_i.$$

In the special case $k = 1$ we notice that $Q(n,1) = -\sum_{i=1}^{n-1} \log \frac{T_i}{T_n}$, a form which brings out the relation with the fixed sample Q-test for testing whether points are distributed at random on a line, already mentioned in § 2. We shall return to this point later.

An important property of the test which will be useful in applications is the relation

$$Q(n,k) = Q(n,1) - Q(k,1).$$

To carry out the test we adopt the rule:

Accept H_0 if $\log R \leqslant \log B$.

Reject H_0 if $\log R \geqslant \log A$.

Continue sampling until either of these inequalities is satisfied.

A and B are the usual constants related to the two kinds of error, α and β, by the equations

$$A \doteqdot (1-\beta)/\alpha, \quad B \doteqdot \beta/(1-\alpha).$$

In practice it is more convenient to rearrange the inequalities so that sampling is continued as long as

$$\frac{-\log B + (n-k)\log(1+a_0)}{a_0} > Q(n,k) > \frac{-\log A + (n-k)\log(1+a_0)}{a_0},$$

the left-hand quantity being the acceptance boundary and the right-hand the rejection boundary. By the method of Cox (1952) it is easy to show that the test terminates with probability one.

It sometimes happens that the sample point $(n, Q(n,k))$ falls near the upper test boundary and is followed by a long delay before the occurrence of the $(n+1)$th event. The question then arises as to whether a decision can be reached before this event occurs. If in the formula for $Q(n+1,k)$ we replace T_{n+1} by the time at which a decision is required, say T' $(T_n < T' < T_{n+1})$ and denote the value of Q so obtained by $Q'(n+1,k)$ we have

$$Q(n+1,k) > Q'(n+1,k).$$

If the point $(n+1, Q'(n+1,k))$ falls outside the boundary so will $(n+1, Q(n+1,k))$, so a decision can be obtained more quickly in this way. No such rule can be used at the lower boundary, nor is one necessary, because a long delay will always move the sample point towards the upper boundary when there is no question of the lower boundary being crossed.

4. PROPERTIES OF THE TEST

(i) *The operating characteristic* (O.C.)

The O.C. function, which is the probability of accepting the null hypothesis when a is the true value, is given by

$$L(a) \doteqdot \frac{A^{h(a)} - 1}{A^{h(a)} - B^{h(a)}},$$

where $h(a)$ is the non-zero root of the equation

$$\int_S \left[\frac{p(v_k \dots v_{n-1} \mid a = a_0)}{p(v_k \dots v_{n-1} \mid a = 0)}\right]^{h(a)} p(v_k, \dots, v_{n-1} \mid a)\, dv_k \dots dv_{n-1} = 1,$$

and S is the region in which the sample point must lie.

Substituting known quantities in this equation, it becomes

$$(1+a_0)^{(n-k)h} \int_S v_{n-1}^{ha_0}\, p(v_k, \dots, v_{n-1} \mid a)\, dv_k \dots dv_{n-1} = 1,$$

or

$$(1+a_0)^{(n-k)h}\, \mathscr{E}(v_{n-1}^{ha_0} \mid a) = 1.$$

Now

$$v_{n-1} = T_k^k T_{k+1} \dots T_{n-1} / T_n^{n-1}.$$

Putting $y_i = T_i/T_n\,(i = k, \ldots, n-1)$, $Y = \mu T_n$ in the joint distribution of the T's, we obtain

$$p(y_1, y_2, \ldots, y_{n-k}) = (1+a)^{n-k}\frac{(n-1)!}{(k-1)!}y_1^{k(a+1)-1}\prod_{i=2}^{n-k} y_i^a.$$

Hence it is found that

$$\mathscr{E}\,(v_{n-1}^{ha_0} \mid a) = \left(1+\frac{ha_0}{1+a}\right)^{-(n-k)},$$

so that the equation for h becomes

$$\left(1+\frac{ha_0}{1+a}\right)^{n-k} = (1+a_0)^{h(n-k)}, \tag{5}$$

which is independent of n and k. Corresponding to any solution of equation (5) with h, a_0 and a there will correspond a second solution with $-h$, b_0 and b where

$$b_0 = -a_0/(1+a_0), \quad b = (a-a_0)/(1+a_0). \tag{6}$$

Table 1. *The value of $L(a)$ for various a*

a	-1	0	a'	a_0	∞
$L\,(a)$ $(a_0<0)$	0	$1-\alpha$	$\dfrac{\log A}{\log A-\log B}$	β	1
$L\,(a)$ $(a_0>0)$	1	$1-\alpha$	$\dfrac{\log A}{\log A-\log B}$	β	0

Then it is easy to show that $L(b) = 1-L(a)$, with α and β interchanged. As h tends to zero, $L(a)$ approaches the value $\log A/(\log A - \log B)$ or $\frac{1}{2}$ if $\alpha = \beta$. Expanding the right-hand side of equation (5) we find that

$$1+h\log(1+a_0)+O(h^2) = 1+ha_0/(1+a),$$

which has a root $h = 0$ if $a = a_0/\log(1+a_0)-1$. Denoting this value of a by a', we have on expansion

$$a' = \tfrac{1}{2}a_0 - \tfrac{1}{12}a_0^2 + O(a_0^3).$$

The five points on the O.C. curve given in Table 1 are sufficient for many purposes when only a rough indication of its shape is required. A more detailed solution can be obtained by solving (5) for other values of h. As a first approximation $h \sim 1-2a/a_0$, more accurate values may be obtained by an iterative procedure or more simply from the intersection of the two graphs

$$y = 1+ha_0/(1+a) \quad \text{and} \quad y = \exp[h\log_e(1+a_0)].$$

Charts could easily be constructed for this purpose if required. As an indication of the extent to which the probability of picking out a departure from randomness falls off when the true value of a is rather less than a_0, the results, all calculated for $a = 0{\cdot}8a_0$, given in Table 2 are perhaps instructive.

(ii) *The average sample number* (A.S.N.)

Let us define

$$z_i = \log\frac{p(v_i \mid v_{i-1}, \ldots, v_k;\ a = a_0)}{p(v_i \mid v_{i-1}, \ldots, v_k;\ a = 0)},$$

which in our case becomes

$$z_i = \log(1+a_0) + a_0(\log v_i - \log v_{i-1}).$$

Wald (1947) showed that, provided the expected value of z_i does not depend on i, the A.S.N. was given by

$$\mathscr{E}_a(n \mid a_0) \simeq \frac{L(a)\log B + (1-L(a))\log A}{\mathscr{E}_a(z_i)},$$

where $\mathscr{E}_a(n \mid a_0)$ and $\mathscr{E}_a(z_i)$ are the expected values of n and z_i when $a = a_0$ is used in the test and the true value is a.

Table 2. *Values of* $(1-L(a))$ *when* $a = 0{\cdot}8a_0$

a_0	a	h	Value of $1-L(a)$ when	
			$\alpha=\beta=0{\cdot}05$	$\alpha=\beta=0{\cdot}01$
$-0{\cdot}5$	$-0{\cdot}4$	$-0{\cdot}516$	0·820	0·915
$-2/7$	$-1{\cdot}6/7$	$-0{\cdot}560$	0·839	0·929
0·4	0·32	$-0{\cdot}633$	0·866	0·948
1·0	0·8	$-0{\cdot}644$	0·876	0·955

By the transformation of the previous section it is easily shown that $\mathscr{E}(-\log v_i) = i/(a+1)$ and hence that

$$\mathscr{E}_a(z_i) = \log(1+a_0) - a_0/(1+a),$$

which is independent of i. At the point $a = a'$ the above expression for the average sample number becomes indeterminate; in this case Wald showed that

$$\mathscr{E}_{a'}(n \mid a_0) \simeq -\log A \log B / \mathscr{E}_{a'}(z_i^2).$$

He also remarked (1947) that the maximum value of $\mathscr{E}_a(n \mid a_0)$ usually occurs at or near this point. By a method similar to that used above, or directly from a theorem of Epstein & Sobel (1954), it may be shown that

$$\mathscr{E}_{a'}(z_i^2) = [\log(1+a_0)]^2,$$

and therefore

$$\mathscr{E}_{a'}(n \mid a_0) \simeq -\log A \log B / [\log(1+a_0)]^2.$$

Numerical values for $\mathscr{E}_a(n \mid a_0)$ are given in Tables 3 and 4. Table 3 shows how the A.S.N. varies with the choice of α, β, a_0 and the true value a. If the table is used with negative values of a_0 and a, then it may easily be shown that

$$\mathscr{E}_b(n \mid b_0) = \mathscr{E}_a(n \mid a_0)$$

with α and β interchanged and the relation between b_0, b and a_0, a as given in equation (6). The row $a = 1$ is of special interest because it represents a linear increase in $\lambda(T)$. Corresponding to this we have $b = 1 + 2b_0$, but this no longer has the same physical significance. The values $a = -1$ and $a = \infty$ represent the most extreme departures from randomness that are obtainable with the chosen form of alternative.

Table 4 gives a more extensive tabulation of $\mathscr{E}_0(n \mid a_0)$ and $\mathscr{E}_{a_0}(n \mid a_0)$ to assist in setting up a test. In the same way as for Table 3 it may be used for negative a_0.

Table 3. *Approximate values of* $\mathscr{E}_a(n \mid a_0)$ *for various* α, β *and* a_0

a \ a_0	β \ α	0·2		0·4		0·6		0·8		1·0		β \ α	b
		0·01	0·05	0·01	0·05	0·01	0·05	0·01	0·05	0·01	0·05		
−1	0·01 or 0·05	0	0	0	0	0	0	0	0	0	0	0·01 or 0·05	−1
−0·25	0·01	54	54	23	23	14	14	10	10	7	7	0·01	$\frac{1}{4}(3b_0-1)$
	0·05	35	35	15	15	9	9	6	6	5	5	0·05	
0	0·01	256	237	72	67	36	33	22	21	16	15	0·01	b_0
	0·05	166	151	47	43	23	21	15	14	11	10	0·05	
a'	0·01	635	409	187	120	96	62	61	41	44	28	0·01	b'
	0·05	409	261	120	77	62	39	41	25	28	18	0·05	
a_0	0·01	289	187	90	58	48	32	32	21	24	16	0·01	0
	0·05	268	170	83	53	45	29	30	19	23	15	0·05	
0·5	0·01	94	61	66	43	—	—	—	—	—	—	0·01	$\frac{1}{2}(1+3b_0)$
	0·05	93	60	64	40	—	—	—	—	—	—	0·05	
1·0	0·01	55	36	34	22	27	18	24	16	24	16	0·01	$1+2b_0$
	0·05	55	36	33	21	26	17	23	15	23	15	0·05	
∞	0·01	25	16	14	9	10	6	8	5	7	4	0·01	∞
	0·05	25	16	14	9	10	6	8	5	7	4	0·05	
		$-\frac{1}{6}$		$-\frac{2}{7}$		$-\frac{3}{8}$		$-\frac{4}{9}$		$-\frac{1}{2}$		b_0	

For $a_0 > 0$, use the top and left-hand marginal scale; for negative a_0, written as b_0, use the bottom and right-hand marginal scale.

(iii) *Saving in sample size effected by the sequential test*

As an alternative to the application of the sequential test just described a succession of fixed sample tests could be used. In this case we should decide in advance to observe a fixed number of events and then to carry out a test for a trend. At the occurrence of the nth event $(n-1)$ observations $T_1, T_2, \ldots, T_{n-1}$ would have been made in the interval $(0, T_n)$. It is easily verified from equation (2) that the variables defined by $y_i = T_i/T_n$ $(i = 1, 2, \ldots, n-1)$ have the same joint distribution as a sample of $(n-1)$ from the population

$$p(y) = (a+1)\, y^a \quad (0 \leqslant y \leqslant 1). \tag{7}$$

Table 4. *The average sample number for the sequential Q-test*

The upper figure is $E_{a_0}(n \mid a_0)$ and the lower $E_0(n \mid a_0)$.

α / β	0·01 / 0·01	0·025 / 0·01	0·05 / 0·01	0·01 / 0·025	0·025 / 0·025	0·05 / 0·025	0·01 / 0·05	0·025 / 0·05	0·05 / 0·05
a_0									
0·175	366	293	237	356	279	229	340	269	216
	329	320	305	263	251	242	213	205	194
0·200	289	231	187	281	220	180	268	212	170
	256	248	237	204	195	188	166	160	151
0·225	235	188	152	228	179	147	218	173	138
	205	199	190	164	156	151	133	128	121
0·25	196	156	127	190	149	122	182	144	116
	169	164	157	135	129	124	109	105	100
0·30	144	115	93	138	110	90	132	106	85
	121	116	112	97	92	89	78	75	71
0·35	111	89	72	108	85	70	103	82	66
	91	89	85	73	70	67	59	57	54
0·40	90	72	58	87	69	56	83	66	53
	72	70	67	58	55	53	47	45	43
0·45	75	60	49	72	57	47	69	55	44
	58	57	54	47	45	43	38	37	35
0·50	63	51	41	62	49	40	59	47	38
	49	47	45	39	37	36	32	30	29
0·60	48	39	32	47	37	31	45	36	29
	36	35	33	29	27	26	23	23	21
0·70	39	31	26	38	30	25	36	29	23
	28	27	26	22	21	21	18	18	17
0·80	32	26	21	32	25	21	30	24	19
	22	22	21	18	17	17	15	14	14
0·90	28	22	18	27	21	18	26	21	17
	18	18	17	15	14	14	12	12	11
1·00	24	20	16	24	19	16	23	18	15
	16	15	15	13	12	12	11	10	10

For negative values of the parameter, take $b_0 = -a_0/(1+a_0)$ and interchange α and β so that we have

$$E_0(n \mid b_0) = E_{a_0}(n \mid a_0), \quad E_{b_0}(n \mid b_0) = E_0(n \mid a_0).$$

Pearson (1938) showed that the test based on the statistic

$$Q_1 = -2\sum_{i=1}^{n-1}\log y_i = -2\sum_{i=1}^{n-1}\log T_i/T_n = 2Q(n,1)$$

was the uniformly most powerful test of the hypothesis $a = 0$ with respect to the class of alternatives defined by (7) with $a > 0$. By making the transformation $u = -2\log y$ in (7) we find that

$$p(u) = (a+1)\,e^{-(a+1)u},$$

from which it follows that Q_1 is distributed as $\chi^2_{2(n-1)}/(1+a)$. A more general form of Q_1 corresponding to $Q(n,k)$ has been given by the author in unpublished work. The following analysis is applicable to the more general case by writing $(n-k+1)$ in place of n.

The purpose of this section is to investigate the saving in sample size achieved by adopting the sequential procedure; this is important because it means that decisions will be reached more quickly using this method.

The number of observations, n, required for the fixed sample test of strength (α, β) is determined by finding the degrees of freedom, $\nu = 2(n-1)$, for which the distribution of χ^2 has the upper 100β % point $= (1+a_0)\times$ its lower 100α % point. If we limit discussion to values of a_0 sufficiently small (say $a_0 \leqslant 0{\cdot}5$), so that ν will be large enough to make use of Fisher's normal approximation to the distribution of χ, i.e. take $\sqrt{(2\chi^2_\nu)}$ to be $N(\sqrt{(2\nu-1)}, 1)$, then for given values of a_0, α and β we have to solve the equation

$$\sqrt{(4n-5)} + X_\beta = (1+a_0)^{\frac{1}{2}}\,(\sqrt{(4n-5)} - X_\alpha).$$

Here X_α and X_β are the appropriate standardized deviates of the normal curve, e.g. $\alpha = \int_{X_\alpha}^{\infty} e^{-\frac{1}{2}x^2}\,dx/\sqrt{(2\pi)}$. Solving the equation for n, we find

$$n = 1 + \tfrac{1}{4}\{1 + (X_\alpha(1+a_0)^{\frac{1}{2}} + X_\beta)^2/(1-(1+a_0)^{\frac{1}{2}})^2\}.$$

Table 5. *Average percentage saving in sample size if sequential test used*

a_0	0·5		0·3		Limit as $a_0 \to 0$		
β \ α	0·05	0·01	0·05	0·01	0·05	0·01	
0·05	57	68	54	56	51	63	H_0 true $a = 0$
0·01	52	63	50	62	47	58	
0·05	43	41	46	43	51	47	H_1 true $a = a_0$
0·01	56	53	59	54	63	58	

Table 5 gives the average saving in sample size, using the sequential test, for various a_0 expressed as a percentage. This is applicable for negative a_0 by taking $b_0 = -a_0/(1+a_0)$, interchanging both α and β, and 'H_0 true' and 'H_1 true'.

The saving in sample size is never less than 40 % over the range of a_0, α and β that are likely to be encountered in practice, in fact it is often considerably more than this.

5. THE TWO-SIDED TEST

Where interest centres on the detection of both increases and decreases, a two-sided version of the test is required.

In this case we test the hypothesis $H_0(a = 0)$ against the alternative $H_1(a = a_0)$ or $H_2(a = -a_1)$; because of the symmetry of the two cases $a = a_0$ and $a = -a_0/(1+a_0)$, it may be convenient to take $a_1 = -a_0/(1+a_0) = b_0$, although this is not necessary, of course.

Two methods have been put forward to deal with this situation. The first uses a likelihood ratio formed by taking the simple average of the ratios required for the separate tests, and the second consists of running two one-sided tests simultaneously. The latter method will be preferred because it requires no additional calculation; it is not difficult to see the relation between the two methods.

Details of how the test is carried out are given in Armitage (1947); the procedure becomes ambiguous if it is possible to accept both H_1 and H_2, but this is only possible for unusual values of a_0, α and β, and even then the probability of such a happening is infinitesimal.

The O.C. function and A.S.N. function have not been obtained for the two-sided test, but in practice our knowledge for the one-sided scheme will be adequate for setting up a test procedure.

6. APPLICATION OF THE TEST

As usual in setting up a sequential test we are faced with the initial problem of deciding on the most appropriate specification of parameter values for the alternative hypotheses H_0 and H_1 (and H_2 also if the two-sided test is used). In the present situation H_0 is clearly the hypothesis of randomness with $a = 0$, but as is frequently the case in other types of problem there is no clear-cut simple alternative H_1. The most appropriate alternative to introduce has to be settled paying regard to the O.C. curve and the A.S.N. associated with different values of the parameter and α and β†.

There is also further difficulty in the present case. The practical significance of changes in the parameter a is less easy to grasp than that of changes in, say, the mean of a normal distribution or the probability p of a binomial. To help in the physical interpretation of a, Table 6 has been prepared giving, for different values of a, the number of events which would occur, on the average, in successive equal periods of time starting from $T = 0$. These results have been standardized by making the expectation in the first period equal to unity. Although these can be obtained from the figures for the first six periods, we have also given at the bottom of the table the expected cumulative totals after 5, 10, ..., 25, 30 periods.

The values in the table are based on the relation

$$\frac{N_1}{N_2} = \left(\frac{t_1}{t_2}\right)^{a+1},$$

where N_1 = expected number occurring in $(0, t_1)$ and N_2 = expected number occurring in $(0, t_2), t_2 > t_1$.

It will be seen that broadly speaking the $\lambda(T)$ law corresponds to a situation where there is a sharp initial change in the expectation with a gradual slowing off. Thus for $a = 0{\cdot}4$, the expectation in the 3rd period is double that in the 1st, but this is not doubled again until

† Were it readily calculable, information regarding the distribution of the decisive sample number as well as its average value would also be valuable.

the 14th period is reached. The way in which this table may be used in combination with Tables 3 and 4 will be illustrated in connexion with the example given below.

The fact that the test may be applied to a sequence of observed times $T_k, T_{k+1}, \ldots$, where $(k-1)$ observations have occurred in the interval $(0, T_k)$, permits a certain amount of variety in its use. The most straightforward application is where the origin for T is taken at the point of time when observations start; here $k = 1$. If a decision is reached at T_{k-1}, and we wish to

Table 6. *Expected number of events under the alternative* $\lambda(T) = \mu(\mu T)^a$ *in successive unit periods expressed as multiples of the expectation for the first period*

Period	$a>0$					$a<0$					Period
	1·0	0·6	0·4	0·3	0·2	$-\frac{1}{6}$	$-\frac{3}{13}$	$-\frac{2}{7}$	$-\frac{3}{8}$	$-\frac{1}{2}$	
1	1·00	1·00	1·00	1·00	1·00	1·00	1·00	1·00	1·00	1·00	1
2	3·00	2·03	1·64	1·46	1·30	0·78	0·70	0·64	0·54	0·41	2
3	5·00	2·77	2·02	1·71	1·44	0·72	0·63	0·55	0·45	0·32	3
4	7·00	3·39	2·30	1·89	1·54	0·68	0·58	0·50	0·39	0·27	4
5	9·00	3·94	2·56	2·04	1·62	0·65	0·54	0·47	0·35	0·24	5
6	11·00	4·45	2·77	2·17	1·69	0·63	0·52	0·44	0·33	0·21	6
7	13·00	4·92	2·96	2·28	1·74	0·61	0·50	0·41	0·31	0·20	7
8	15·00	5·36	3·13	2·38	1·80	0·59	0·48	0·40	0·30	0·18	8
9	17·00	5·77	3·29	2·47	1·84	0·58	0·47	0·39	0·28	0·17	9
10	19·00	6·18	3·45	2·55	1·88	0·57	0·46	0·38	0·27	0·16	10
15	29·00	7·96	4·08	2·90	2·05	0·53	0·42	0·33	0·23	0·13	15
20	39·00	9·51	4·59	3·17	2·17	0·51	0·39	0·31	0·21	0·11	20
25	49·00	10·91	5·04	3·39	2·27	0·49	0·37	0·29	0·19	0·10	25
30	59·00	12·19	5·42	3·59	2·36	0·48	0·35	0·27	0·18	0·09	30
	Expected cumulative totals										
1–5	25·00	13·13	9·52	8·10	6·90	3·82	3·45	3·16	2·73	2·24	1–5
1–10	100·00	39·81	25·12	19·95	15·85	6·81	5·88	5·18	4·22	3·16	1–10
1–15	225·00	76·16	44·31	33·80	25·78	9·55	8·03	6·92	5·43	3·87	1–15
1–20	400·00	120·68	66·29	49·13	36·41	12·14	10·02	8·50	6·50	4·47	1–20
1–25	625·00	172·47	90·60	65·66	47·59	14·62	11·89	9·97	7·48	5·00	1–25
1–30	900·00	230·88	116·94	83·23	59·23	17·02	13·68	11·35	8·38	5·48	1–30

continue looking out for changes in $\lambda(T)$, we may now start again with a new sequence of times $T_i^* = T_{k+i-1} - T_{k-1}$ $(i = 1, 2, \ldots)$. This procedure would clearly be appropriate if, for example, the events were machine failures and the machine was reset when an increase in the failure rate had been established.

On the other hand, if a decision in favour of H_0 is reached after $(k_1 - 1)$ observations it is possible to continue observation with T measured from the initial origin. The test statistic to be compared with the boundaries given in § 3 would now be

$$Q(n, k_1) = Q(n, 1) - Q(k_1, 1).$$

It is important to notice that the inference made at the second decision in this case applies to the *whole period from the initial origin* and not to the period from the last decision. As long

as decisions of the same kind are reached it would seem that the process can be continued, using $Q(n, k_s) = Q(n, 1) - Q(k_s, 1)$ after the sth decision. If, however, there is a change in decision from H_0 to H_1 or vice versa, justification for the further use of the original origin appears more doubtful. It is clear that here, as in all sequential applications, if a change in the probability law occurs during the course of observation, the strict theoretical basis of the test breaks down and its utility must to a large extent be based on empirical study. Some further investigation of this kind is in hand in the present case.

7. A NUMERICAL EXAMPLE

We have taken the data given by Maguire *et al.* (1952) for explosions in coal-mines in Great Britain since 1875 involving the loss of ten lives or more. These authors gave the intervals in days between successive accidents from which values of T with origin at 6 December 1875 can readily be found. Fig 1 shows a plot of the observations.

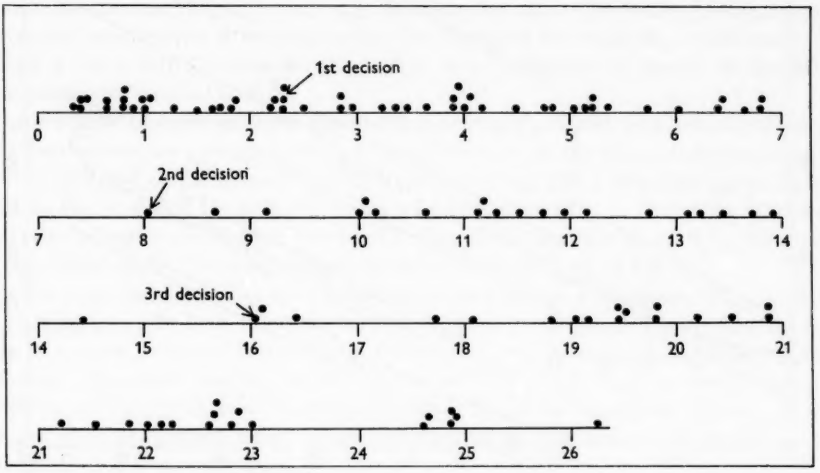


Fig. 1. Time plot of mine accident data. Time scale: one unit = 1000 days. Origin at 6 December, 1875.

Let us suppose that, starting from the beginning of the record, we are on the look-out for evidence of a *reduction* in accident rate. This means that a_0 must be taken negative; for clearness in reference to the tables we shall denote this value by b_0. In a situation of this kind we should like to reach a provisional opinion as to what was happening fairly rapidly and there would be no object in playing for safety using very small values of α and β. Let us therefore take the largest values of α and β given in Tables 3 and 4, namely, $\alpha = \beta = 0{\cdot}05$. Then, using Table 3, let us examine the consequences of choosing particular values for b_0.

(i) If we adjust the test so as to be efficient in picking out a small value of b_0, say $-\frac{1}{6}$, we must expect to wait a very long time for an answer. It will be seen that $\mathscr{E}_0(n \mid b_0) = 170$ and $\mathscr{E}_{b_0}(n \mid b_0) = 151$. Supposing an accident rate of 3 per annum in the year 1875, this means that using the test so adjusted we might expect to wait 57 years before establishing that $\lambda(T)$

was remaining constant. The reason for this is that a change in $\lambda(T)$ represented by the expectations in the column headed $b = -\frac{1}{6}$ in Table 5 would be very hard to distinguish from a random situation because, after the initial drop in frequency (which may be obscured by sampling fluctuations) there is only a very gradual falling off. For example, as Table 5 shows, with $b = -\frac{1}{6}$ the expectation in the 30th period is 26 % below that in the 5th period while for $b = -\frac{3}{8}$ it is 49 % less.

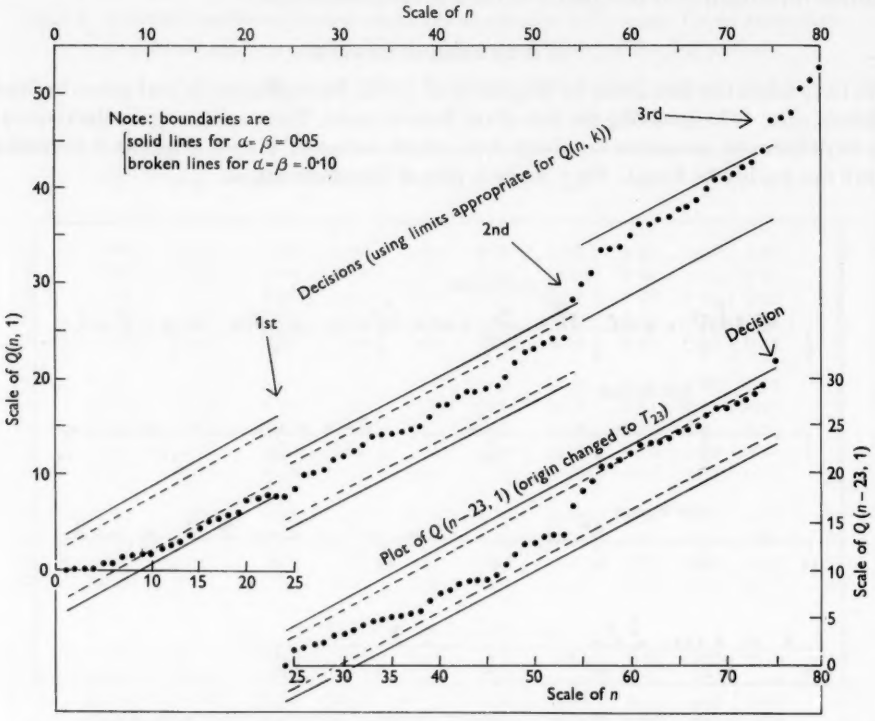

Scale of n
Note: boundaries are
solid lines for α = β = 0·05
broken lines for α = β = 0·10
Decisions (using limits appropriate for Q(n, k))
1st
2nd
3rd
Decision
Plot of Q(n−23, 1) (origin changed to T23)
Scale of Q(n, 1)
Scale of Q(n − 23, 1)
Scale of n

Fig. 2. Mine accident data; application of sequential tests with $a_0(=b_0)=\frac{1}{3}$.

(ii) If $b_0 = -\frac{1}{6}$ and the true decrease is more marked, say $b = -\frac{3}{8}$, then we see from Table 3 that this change would be detected, on average, after 35 observations.† This is more satisfactory, but we note that if the test had been adjusted to be most efficient for $b = -\frac{3}{8}$, i.e. by making $b_0 = -\frac{3}{8}$, then on the average only 21 observations instead of 35 would have been needed to reach a decision.

(iii) Keeping $b_0 = -\frac{1}{6}$ we should expect to detect a smaller decrease of $b = b' = -\frac{1}{12}$ with probability $\frac{1}{2}$, but this, on average, would take 261 observations. To sum up, with $b_0 = -\frac{1}{6}$, we shall stand a good chance of detecting any decrease of importance, but we may have to wait a very long time to do so.

(iv) If we go to the other extreme and take $b_0 = -\frac{4}{9}$, Table 3 shows that a decision will be reached fairly quickly whatever the true value of b. A decrease with $b = \frac{1}{4}(3b_0 - 1) = -\frac{7}{12}$

† This figure comes from the section of the table with $b_0 = -\frac{1}{6}$, $b = \frac{1}{4}(3b_0 - 1) = -\frac{3}{8}$.

will be reached almost as quickly as if b_0 had been chosen equal to $-\frac{7}{12}$; however, though decisions will be reached quickly if $b < b_0$, say in the neighbourhood of $-\frac{2}{3}$, half of them will be incorrect, i.e. half will be in favour of H_0 or randomness.

These considerations suggest that an intermediate value of b_0, say $-\frac{1}{3}$ (corresponding to the positive value of $a_0 = 0{\cdot}5$) will be most appropriate. If the series is random we should reach a decision for H_0 on the average after 43 observations, while if the expectation is decreasing with $a_0 = -\frac{1}{3}$, after 29 observations.

If we are prepared to take greater risks of wrong decisions, e.g. make $\alpha = \beta = 0{\cdot}10$, the process could, of course, be speeded up.

Fig. 2 shows the values of Q plotted against n, taking $a_0 = -\frac{1}{3}$. This diagram should be studied in conjunction with Fig. 1. We start on the left with values of

$$Q(n, 1) = (n-1)\log_{10} T_n - \sum_{i=1}^{n-1} \log_{10} T_i,$$

and draw the boundary lines for $\alpha = \beta = 0{\cdot}05$ (solid rules) at $(n-1) \times 0{\cdot}528 \pm 3{\cdot}836$. A decision in favour of randomness is reached at the 23rd accident for which $T_{23} = 2326$ days. Had we used $\alpha = \beta = 0{\cdot}10$ (boundaries, the broken lines) a decision in favour of randomness would have been reached at T_{10}.

If it is wished to continue the sequential test beyond T_{23} (about the beginning of the year 1883), two courses are now open. We may keep the origin at the original starting point and calculate $Q(n, 24)$; this is effectively done by still plotting $Q(n, 1)$ but shifting the limits as shown in Fig. 2, so that the new origin coincides with point $Q(k_1, 1)$. Following this plan we reach two successive decisions in favour of a drop in the accident rate: at T_{54} (8042 days), and T_{75} (16051 days). These results are shown in the upper part of Fig. 2.

On the other hand, we may start again with a new origin at T_{23} taking $T^*_{n-23} = T_n - T_{23}$ and plotting $Q(n-23, 1)$ as shown in the lower part of the diagram. In this case, for a long while it appears as though a decision in favour of H_0 will be reached, but after T_{54}, or the 31st observation since the first decision, the trend is reversed. Had we used $\alpha = \beta = 0{\cdot}10$ a decision in favour of H_1 would have been made at T_{57} (10,024 days), but with $\alpha = \beta = 0{\cdot}05$ it is not until T_{75} (16,051 days) is reached that we should come down in favour of an established drop in accident rate.

It is impossible to draw any general conclusions from this simple example, particularly since, when a drop occurs, it cannot be described in terms of a single value of the parameter b. The appearance of the spots in Fig. 1 suggests the existence of several fluctuations in rate. We think, however, that the illustration throws some light on the working of the test. It also suggests points for further investigation. We may ask whether the second decision is reached more quickly when the origin is *not* changed after the first decision,

(*a*) because the fall-off in λ is better represented by $\mu(\mu T)^a$ than by $\mu\{\mu(T - T_{23})\}^a$,

(*b*) or because the high initial density before T_{23} still plays its part through the value of k in enhancing the term $-(k-1)\log T_k$ in the expression for $Q(n, k)$?

Another point which requires further investigation is the behaviour of the test under alternative trends having different mathematical forms from the one on which the sequential Q test is based. A beginning has been made by applying the test to two artificial series generated by the law $\lambda(T) = \frac{1}{4}(1 + \frac{1}{4}\mu T)$. Details of the results are given in Bartholomew (1955).

8. Summary

A method has been given for testing whether a sequence of events is occurring at random in time or space when the alternative is a trend. The test is derived on the assumption that the alternative can be represented by a function of the form $\lambda(T) = \mu(\mu T)^a$ for $-1 < a < \infty$. The theory of sequential analysis is used to derive the test together with its operating characteristic and average sample number functions. The application of the test under various circumstances has been discussed from the theoretical standpoint and illustrated on mining accident data.

My thanks are due to Dr N. L. Johnson, who supervised this work, and to Prof. E. S. Pearson for much valuable help in preparing the work for publication; also to the Department of Scientific and Industrial Research for the award of a maintenance grant.

REFERENCES

Armitage, P. (1947). *J. R. Statist. Soc.* B, **9**, 250.

Bartholomew, D. J. (1955). Ph.D. Thesis, University of London.

Cox, D. R. (1952). *Proc. Camb. Phil. Soc.* **48**, 290.

Epstein, B. & Sobel, M. (1954). *Ann. Math. Statist.* **25**, 373.

Maguire, B. A., Pearson, E. S. & Wynn, A. H. A. (1952). *Biometrika*, **39**, 168.

Pearson, E. S. (1938). *Biometrika*, **30**, 134.

Wald, A. (1947). *Sequential Analysis*. New York: John Wiley and Sons Inc.; London: Chapman and Hall Ltd.

ON THE MOMENTS OF THE MAXIMUM OF PARTIAL SUMS OF A FINITE NUMBER OF INDEPENDENT NORMAL VARIATES

By A. A. ANIS

Chelsea Polytechnic, London

SUMMARY. The paper is concerned with the maximum U_n of partial sums X_1, X_1+X_2, ..., $X_1+X_2+\ldots+X_n$ of n independent standard normal variates.

The distribution of U_n is of interest in the theory of storage. Suppose we have a reservoir of infinite capacity, which receives every year a random input, from rivers, etc., whose distribution is normal $(\mu, 1)$, and releases, for civil purposes, the mean discharge μ. The probability that, starting with an initial water level x, the reservoir will not run dry in the following n years is given by the distribution function $F_n(x)$ of U_n.

The first and second moments of the variate U_n have been previously obtained (Anis & Lloyd, 1953; Anis, 1955). Each of these moments was studied on its own merits, and no systematic method of attack was seen at that time. In the present paper, a method for obtaining all the moments is discussed. A recurrence relation is obtained which makes possible their numerical evaluation.

1. Statement of the problem

Consider n independent standard normal variates $X_1, X_2, \ldots, X_n$, and their partial sums

$$S_r = X_1+\ldots+X_r \quad (r = 1, 2, \ldots, n).$$

Let

$$U_n = \operatorname*{Max}_r \{S_r\}$$

denote the maximum of these partial sums. Our problem is to obtain the moments of U_n. We shall always use the symbols $\phi(x)$, $\Phi(x)$ to denote respectively the frequency function and the distribution function of a standard normal variate. We shall also use $F_n(x), f_n(x)$ to denote respectively the distribution function and the frequency functions of U_n, i.e.

$$F_n(x) = \Pr(U_n \leqslant x), \quad f_n(x) = F_n'(x).$$

We have

$$F_n(y) = \int_K (n) \int \prod_1^n \phi(x_i)\, dx_i, \tag{1·1}$$

where the region of integration K is defined by

$$K: \quad \sum_1^r x_i \leqslant y \quad (r = 1, 2, \ldots, n).$$

2. The moments as a linear form of the $M_j(n+1)$

It may be deduced from (1·1) that

$$F_{n+1}(x) = \int_0^\infty F_n(t)\, \phi(x-t)\, dt, \tag{2·1}$$

and hence that

$$f_{n+1}(x) = F_n(0)\, \phi(x) + \int_0^\infty f_n(t)\, \phi(x-t)\, dt. \tag{2·2}$$

Also it is well known that

$$\phi(x-t) = \phi(x) \sum_{j=0}^\infty t^j H_j(x)/j!, \tag{2·3}$$

where $H_j(x)$ is the Hermite polynomial of degree j. On substituting from (2·3) into (2·2) we get

$$f_{n+1}(x) = \phi(x) \sum_{j=0}^\infty H_j(x)\, M_j(n+1)/j!, \tag{2·4}$$

where $$M_j(n+1) = \int_0^\infty t^j f_n(t)\,dt \quad (j \neq 0)$$

and $$M_0(n+1) = 1.$$ (2·5)

Multiplying both sides of (2·4) by $H_j(x)$ and integrating over the range $(-\infty, \infty)$, we obtain

$$M_j(n+1) = \int_{-\infty}^{\infty} H_j(x) f_{n+1}(x)\,dx. \tag{2·6}$$

Using the well-known explicit forms for the $H_j(x)$, we easily obtain the following relations:

$$\left.\begin{aligned} \mu_1'(n+1) &= M_1(n+1), \\ \mu_2'(n+1) &= 1 + M_2(n+1), \\ \mu_3'(n+1) &= 3M_1(n+1) + M_3(n+1), \text{ ... etc.,} \end{aligned}\right\} \tag{2·7}$$

where $$\mu_r'(n+1) = \int_{-\infty}^{\infty} x^r f_{n+1}(x)\,dx.$$ (2·8)

Hence the problem of evaluating the moments is reduced to that of evaluating the functions $M_j(n+1)$.

3. $M_j(n+1)$ AS A LINEAR FORM IN THE $F_r(0)$

On applying equation (2·2) to reduce $f_n(t)$ to $f_{n-1}(t)$, we obtain

$$M_j(n+1) = F_n(0)\int_0^\infty t^j \phi(t)\,dt + \int_0^\infty \int_0^\infty t^j f_{n-1}(y)\,\phi(y-t)\,dy\,dt.$$

Continuing this process of reduction from f_{n-1} to f_{n-2}, ..., etc., we finally get

$$M_j(n+1) = \sum_{r=1}^{n} a(r,j)\,F_{n-r}(0),$$ (3·1)

where $$\left.\begin{aligned} a(r,j) &= \int_0^\infty (r) \int_0^\infty y_1^j \phi(y_1 - y_2)\,\phi(y_2 - y_3) \cdots \phi(y_{r-1} - y_r)\,\phi(y_r) \prod_1^r dy_i \quad (r \geqslant 2), \\ a(1,j) &= \int_0^\infty y_1^j \phi(y_1)\,dy_1. \end{aligned}\right\} \tag{3·2}$$

It may be appropriate to recall at this stage an important lemma proved in a previous paper (Anis, 1955) and which we need in the sequel. This is

$$\sum_{r=0}^{\infty} F_r(0)\,t^r = (1-t)^{-1}.$$ (3·3)

4. A RECURRENCE RELATION FOR THE $a(r,j)$

In an earlier paper (Anis & Lloyd, 1953) the integral

$$b_s = \int_0^\infty (s-1) \int_0^\infty \phi(y_1)\,\phi(y_1 - y_2) \cdots \phi(y_{s-2} - y_{s-1})\,\phi(y_{s-1}) \prod_1^{s-1} dy_i = (2\pi s^3)^{-\frac{1}{2}}$$ (4·1)

was discussed. Here we shall show the whole results depend on this integral. We make use in (3·2) of the identity

$$y_1 = \sum_{s=1}^{r} (r-s+1)\,(2y_s - y_{s-1} - y_{s+1})$$ (4·2)

(with the convention that $y_0 = y_1$, $y_{r+1} = 0$); hence

$$a(r,j) = \sum_{s=1}^{r} \int_0^\infty (r) \int_0^\infty (r-s+1)\,(2y_s - y_{s-1} - y_{s+1})\, y_1^{j-1} \phi(y_1 - y_2) \dots \phi(y_{r-1} - y_r)\, \phi(y_r) \prod_1^r dy_i. \tag{4·3}$$

The first term of this sum may be integrated by parts and has the value

$$r(j-1)\,a(r,j-2) \quad (j \geqslant 2).$$

The other terms may also be integrated by parts; the presence of the terms $(2y_s - y_{s-1} - y_{s+1})$ has the effect of dividing every integral into two multiplied parts, one of the form $a(r,j)$ for some r, j and the other of the form b_s for some s. This leads us directly to the recurrence relation

$$a(r,j) = r(j-1)\,a(r,j-2) + \sum_{s=1}^{r-1} (r-s)\, b_{r-s}\, a(s,j-1). \tag{4·4}$$

If $j = 1$, we get

$$a(r,1) = \sum_{s=1}^{r} s b_s F_{r-s}(0); \tag{4·5}$$

the introduction of $F_s(0)$ in this last relation is due to the fact that

$$F_s(0) = \int_0^\infty (s) \int_0^\infty \phi(y_1 - y_2)\, \phi(y_2 - y_3) \dots \phi(y_{s-1} - y_s)\, \phi(y_s) \prod_1^s dy_i, \tag{4·6}$$

i.e. $F_s(0) = a(s,0)$. If we define c_r as

$$c_r = r b_r = (2\pi r)^{-\frac{1}{2}}, \tag{4·7}$$

then (4·4) and (4·5) may be written in the form

$$\left.\begin{aligned} a(r,j) &= r(j-1)\,a(r,j-2) + \sum_{s=1}^{r-1} c_{r-s}\, a(s,j-1) \quad (j \geqslant 2), \\ a(r,1) &= \sum_{s=1}^{r} c_s F_{r-s}(0). \end{aligned}\right\} \tag{4·8}$$

We now have the position that the required moments $\mu'_r(n)$ of the original variate U_n are known (2·7) in terms of the functions $M_s(n)$, which are themselves known (3·1) in terms of the functions $a(r,j)$; and we have a recurrence relation for the $a(r,j)$ involving only the known coefficients c_r. We shall not require the values of these functions $a(r,j)$; in the next section we replace the recurrence relation for $a(r,j)$ by a difference-differential equation for their generating functions, and show that from these we may construct a generating function for the $M_s(n)$.

5. A DIFFERENCE-DIFFERENTIAL EQUATION OF THE GENERATING FUNCTION $a(r,j)$

Let us define the generating function $Q_j(t)$ as follows:

$$\left.\begin{aligned} Q_j(t) &= \sum_{r=1}^{\infty} a(r,j)\, t^r \quad (j \geqslant 1), \\ Q_0(t) &= \sum_{r=0}^{\infty} a(r,0)\, t^r = \sum_{r=0}^{\infty} F_r(0)\, t^r = (1-t)^{-\frac{1}{2}} \end{aligned}\right\} \tag{5·1}$$

(using the lemma of (3·3)). From (4·8) we know that

$$\left.\begin{aligned} a(1,j) &= (j-1)\,a(1,j-2), \\ a(2,j) &= 2(j-1)\,a(2,j-2) + a(1,j-1)\,c_1, \\ a(3,j) &= 3(j-1)\,a(3,j-2) + a(1,j-1)\,c_2 + a(2,j-1)\,c_1, \text{ etc.} \end{aligned}\right\} \tag{5·2}$$

If we multiply both sides of the first equation by t, the second by t^2, and so on, and add the terms $a(s, j-2)$ vertically but those of $a(s, j-1)$ diagonally, we get

$$Q_j(t) = (j-1)\, tQ'_{j-2}(t) + \alpha(t)\, Q_{j-1}(t) \quad (j \geqslant 2), \tag{5·3}$$

where

$$\alpha(t) = \sum_{r=1}^{\infty} c_r t^r. \tag{5·4}$$

$\alpha(t)$ is absolutely convergent for all values of t in the range $|t| < 1$. If $j = 1$, then the second relation of (4·8) would lead us to

$$Q_1(t) = \alpha(t)\, Q_0(t). \tag{5·5}$$

We may observe with the aid of (3·1) that

$$\begin{aligned} Q_0(t)\, Q_j(t) &= (1-t)^{-\frac{1}{2}} Q_j(t) \\ &= \sum_{n=1}^{\infty} \left\{ \sum_{r=1}^{n} a(r, j)\, F_{n-r}(0) \right\} t^n, \end{aligned}$$

or

$$Q_0(t)\, Q_j(t) = \sum_{n=1}^{\infty} M_j(n+1)\, t^n, \tag{5·6}$$

i.e. $Q_0(t)\, Q_j(t)$ is the generating function of $M_j(n+1)$. We have earlier reduced the question of evaluating the moments to that of evaluating $M_j(n+1)$. Equation (5·6) reduces this last question to that of obtaining $Q_0(t)\, Q_j(t)$.

It may be appropriate here to work out, as examples, the explicit expressions of M_1, M_2, M_3.

The value of $Q_1(t)$ is given by (5·5), $Q_2(t)$ and $Q_3(t)$ could be obtained directly from (5·3) as functions of $\alpha(t)$, $Q_0(t)$ and their derivatives. By this method we get

$$\begin{aligned} Q_0(t)\, Q_1(t) &= (1-t)^{-1} \alpha(t), \\ Q_0(t)\, Q_2(t) &= \tfrac{1}{2} t (1-t)^{-2} + (1-t)^{-1} \alpha^2(t), \\ Q_0(t)\, Q_3(t) &= \tfrac{3}{2} (1-t)^{-2} \alpha(t) + 2t(1-t)^{-1} \alpha'(t) + (1-t)^{-1} \alpha^3(t). \end{aligned}$$

Equating the coefficient of t^n on both sides of these equations we get

$$M_1(n+1) = (2\pi)^{-\frac{1}{2}} \sum_{r=1}^{n} r^{-\frac{1}{2}}.$$

$$M_2(n+1) = \tfrac{1}{2} n + (2\pi)^{-1} \sum_{r=1}^{n-1} \sum_{s=1}^{r} \{s(r-s+1)\}^{-\frac{1}{2}},$$

$$M_3(n+1) = \tfrac{3}{2} n (2\pi)^{-\frac{1}{2}} \sum_{r=1}^{n} r^{-\frac{1}{2}} + \tfrac{1}{2} (2\pi)^{-\frac{1}{2}} \sum_{r=1}^{n} r^{\frac{1}{2}} + (2\pi)^{-\frac{3}{2}} \sum_{r=1}^{n-2} \sum_{s=1}^{r} \sum_{k=1}^{s} \{k(s-k+1)\,(r-s+1)\}^{-\frac{1}{2}}. \tag{5·7}$$

The values of M_1, M_2 and M_3 were computed from these formulae. The third term of M_3 was computed using the recurrence relation

$$p_3(n) = \sum_{r=1}^{n} r^{-\frac{1}{2}} p_2(n-r+1), \tag{5·8}$$

where

$$p_2(n) = \sum_{r=1}^{n} \sum_{s=1}^{r} \{s(r-s+1)\}^{-\frac{1}{2}} \tag{5·9}$$

and

$$p_3(n) = \sum_{r=1}^{n} \sum_{s=1}^{r} \sum_{k=1}^{s} \{k(s-k+1)\,(r-s+1)\}^{-\frac{1}{2}}. \tag{5·10}$$

Since the values of $p_2(r)$ are available from the computation of M_2, this recurrence relation facilitates the computation of $p_3(r)$ (and hence M_3) considerably. Unfortunately, this procedure could not be continued much further since the formulae rapidly become unmanageable.

In the following section we show how to obtain $M_r(n+1)$ for $r>3$, and hence the moments of U_n, from recurrence relations which depend on the values of M_1, M_2 and M_3 obtained above.

We may also mention that an explicit solution may be found for the difference-differential equation (5·3). Details will be published elsewhere.

6. A RECURRENCE RELATION IN $M_j(n+1)$

We have shown that

$$Q_0(t)\,Q_j(t) = \sum_{n=1}^{\infty} M_j(n+1)\,t^n.$$

Hence

$$Q_j(t) = (1-t)^{\frac{1}{2}} \sum_{n=1}^{\infty} M_j(n+1)\,t^n, \tag{6·1}$$

and

$$Q_j'(t) = -\tfrac{1}{2}(1-t)^{-\frac{1}{2}} \sum_{n=1}^{\infty} M_j(n+1)\,t^n + (1-t)^{\frac{1}{2}} \sum_{n=1}^{\infty} nM_j(n+1)\,t^{n-1}. \tag{6·2}$$

On substituting from (6·1), (6·2) into (5·3), and equating the coefficient of t^n on both sides of the equation we obtain the recurrence relation

$$M_j(n+1) = \sum_{r=1}^{n-1} c_r M_{j-1}(n-r+1) + (j-1)\,nM_{j-2}(n+1) - \tfrac{1}{2}(j-1) \sum_{r=1}^{n-1} M_{j-2}(r+1). \tag{6·3}$$

This recurrence relation finally gives us a practical method for the numerical computation of the moments. The computation of $M_4(n+1)$ was done by using (6·3), taking for M_1, M_2 and M_3 the values given in § 5. From these it is a very small step to the first four moments of U_n. The following table gives for $n = 2(1)15$ the mean value of U_n; the moments about the mean, μ_2, μ_3, μ_4; and the moment ratios $\gamma_1 = \mu_3/\mu_2^{\frac{3}{2}}$, $\gamma_2 = \mu_4/\mu_2^2 - 3$. It will be seen that as n increases, the distribution becomes increasingly asymmetrical and leptokurtic.

Short table of the first four moments of γ_1, γ_2

n	μ_1'	μ_2	μ_3	μ_4	γ_1	γ_2
2	0·3989	1·3408	0·3265	5·6733	0·2103	0·1556
3	0·6810	1·6953	0·7881	9·3981	0·3570	0·2698
4	0·9114	2·0536	1·3540	14·1469	0·4601	0·3544
5	1·1108	2·4136	2·0075	19·9160	0·5354	0·4186
6	1·2893	2·7745	2·7385	26·7048	0·5926	0·4691
7	1·4521	3·1360	3·5392	34·5131	0·6373	0·5094
8	1·6029	3·4978	4·4042	43·3413	0·6732	0·5425
9	1·7440	3·8599	5·3289	53·1895	0·7027	0·5700
10	1·8769	4·2222	6·3098	64·0580	0·7273	0·5932
11	2·0031	4·5847	7·3438	75·9468	0·7481	0·6132
12	2·1234	4·9473	8·4282	88·8562	0·7660	0·6304
13	2·2385	5·3100	9·5609	102·7862	0·7814	0·6455
14	2·3492	5·6727	10·7400	117·7369	0·7949	0·6588
15	2·4558	6·03551	11·9634	133·7086	0·8068	0·6705

7. The limiting values of γ_1, γ_2

From the linear relations between the M_j's and the moments of U_n about the origin, it is easy to obtain the following relations:

$$\left.\begin{aligned} \mu_2 &= M_2 - M_1^2 + 1, \\ \mu_3 &= M_3 - 3M_1M_2 + 2M_1^3, \\ \mu_4 &= M_4 - 4M_1M_3 + 6M_1^2M_2 - 3M_1^4 + 6M_2 - 6M_1^2 + 3. \end{aligned}\right\} \tag{7·1}$$

Hence
$$\left.\begin{aligned} \gamma_1 &= (M_3 - 3M_1M_2 + 2M_1^3)/(M_2 - M_1^2 + 1)^{\frac{3}{2}}, \\ \gamma_2 &= (M_4 - 4M_1M_3 + 6M_1^2M_2 - 3M_1^4 + 6M_2 - 6M_1^2 + 3)/(M_2 - M_1^2 + 1)^2 - 3. \end{aligned}\right\} \tag{7·2}$$

Now it may be proved that

$$\lim_{n\to\infty} n^{-\frac{1}{2}j} M_j = 2^{\frac{1}{2}j}\Gamma\left(\frac{j+1}{2}\right)\Big/\Gamma(\tfrac{1}{2}). \tag{7·3}$$

This may be done by induction, using the recurrence relation (6·3) of the M_j's, and substituting for M_{j-1}, M_{j-2} their asymptotic values obtained from (7·3) for sufficiently large n. The summation sign in the recurrence relation would be replaced by an integral sign in the course of the proof. Since (7·3) is true for $j = 1, 2$ (Anis & Lloyd, 1953; Anis, 1955) it is in general true.

Computing these limits for $j = 3, 4$ from (7·3) we may proceed easily to obtain the limiting values of γ_1, γ_2 when n tends to infinity. These are

$$\gamma_1 = \sqrt{\frac{2}{\pi}}\left(\frac{4}{\pi} - 1\right)\Big/\left(1 - \frac{2}{\pi}\right)^{\frac{3}{2}} \approx 0{\cdot}992,$$

$$\gamma_2 = \frac{8}{\pi}\left(1 - \frac{3}{\pi}\right)\Big/\left(1 - \frac{2}{\pi}\right)^2 \approx 0{\cdot}875.$$

I am grateful to Prof. E. S. Pearson for drawing my attention to the question of the limiting values of γ_1 and γ_2. I am also indebted to Mr S. Michaelson (Imperial College) for his part in the computation.

REFERENCES

Anis, A. A. (1955). The variance of the maximum of partial sums of a finite number of independent normal variates. *Biometrika*, **42**, 96.

Anis, A. A. & Lloyd, E. H. (1953). On the range of partial sums of a finite number of independent normal variates. *Biometrika*, **40**, 35.

ON THE APPLICATION TO STATISTICS OF AN ELEMENTARY THEOREM IN PROBABILITY

By H. A. DAVID

Commonwealth Scientific and Industrial Research Organization, Sydney, Australia, and University of Melbourne

1. Introduction

Let $P_{ij\ldots k}$ denote the joint probability of $r\,(\leqslant n)$ events $A_i, A_j, \ldots, A_k$; and S_r the sum of the $\binom{n}{r}$ P's with r different subscripts. Then the probability $p_{m,n}$ of the realization of at least m events out of n is given by (see, for example, Feller, 1950, p. 74)

$$p_{m,n} = \sum_{l=0}^{n-m} (-1)^l \binom{m+l-1}{m-1} S_{m+l}. \tag{1}$$

We shall be concerned with various statistical applications of this theorem. Let $y_1, y_2, \ldots, y_n$ be a set of random variables not necessarily independent or identically distributed, and $y_{(1)}, y_{(2)}, \ldots, y_{(n)}$ the same variates arranged in descending order of magnitude. We may now write $P_{ij\ldots k}(Y)$ for the joint probability of the r events $y_i > Y, y_j > Y, \ldots, y_k > Y$. Correspondingly, $p_{m,n}(Y)$ will be the probability that $y_{(m)}$ exceeds Y. In the special case when the joint distribution of the y's is a symmetric function of the y's we have

$$S_r(Y) = \binom{n}{r} P_{12\ldots r}(Y),$$

and (1) takes the form

$$p_{m,n}(Y) = \Pr(y_{(m)} > Y) = \sum_{l=0}^{n-m} (-1)^l \binom{n}{m+l} \binom{m+l-1}{m-1} P_{12\ldots(m+l)}(Y). \tag{2}$$

In particular, for the important case $m = 1$ we have simply

$$\Pr(y_{(1)} > Y) = \sum_{l=1}^{n} (-1)^{l-1} \binom{n}{l} P_{12\ldots l}(Y). \tag{3}$$

Equation (3) has in fact been used by Cochran (1941) to determine the distribution of the ratio of the largest to the sum of a group of sample variances all following a χ^2 law with the same number of degrees of freedom. However, (3) and its generalizations are of much wider applicability as many important statistics are expressible as maxima. Examples are the maximum F-ratio (the largest of a set of F-ratios with a common denominator); the range, the extreme deviate from the sample mean, and their studentized forms. The method is particularly convenient for the determination of upper percentage points, on the assumption of a normal parent population.* In a number of cases it permits also the evaluation of a power function, but the interpretation of this function needs a little care in the present context.

It may be mentioned here that since this paper was first submitted for publication an approach very close to ours has been employed by Halperin, Greenhouse, Cornfield &

* Similar remarks apply, of course, to lower percentage points of statistics expressible as minima (cf. Hartley, 1938).

Zalokar (1955) to obtain upper percentage points for the 'studentized maximum absolute deviate in normal samples'. Some of the criteria listed above will now be considered in more detail.

2. The extreme deviate from the sample mean

Let $x_t\,(t = 1, \ldots, n)$ be n independent normal variates with unit standard deviation. Applying equation (3) to $y_t = x_t - \bar{x}$ we have

$$\Pr(x_{\max.} - \bar{x} > Y) = \binom{n}{1}\Pr(x_1 - \bar{x} > Y) - \binom{n}{2}\Pr(x_1 - \bar{x} > Y, x_2 - \bar{x} > Y) + \ldots$$

$$= T_1(Y) - T_2(Y) + \ldots + (-1)^{n-1}T_n(Y) \quad \text{(say)}. \tag{4}$$

Clearly, for reasonably large values of Y a first approximation to the left-hand side is provided by T_1, with $T_2, T_3, \ldots$ as successive correction terms.* This first approximation, namely

$$\Pr(x_{\max.} - \bar{x} > Y) \doteqdot n\int_{Y\sqrt{[n/(n-1)]}}^{\infty} \frac{e^{-\frac{1}{2}z^2}}{\sqrt{(2\pi)}}\,dz,$$

was suggested by McKay (1935), but its remarkable accuracy for the determination of the usual range of upper percentage points was established only when the exact values were tabulated for $n \leqslant 25$ (Nair, 1948*b*; Grubbs, 1950). The position is easily appreciated from (4) in which T_2 may be expected to be small for large Y since $x_1 - \bar{x}$, $x_2 - \bar{x}$ are negatively correlated. T_2 can, of course, be evaluated from tables of the volume under the normal bivariate surface. However, without doing this it is possible to obtain a simple gauge for the accuracy of $Y_1(\alpha)$, the first term approximation to the upper $100\alpha\%$ point Y_α of $x_{\max.} - \bar{x}$.

By Bonferroni's inequalities (Feller, 1950)

$$T_1 - T_2 < \Pr(x_{\max.} - \bar{x} > Y) < T_1.$$

But for $Y = Y_1$ we have $\quad T_1 = \alpha$

and $\quad T_2 < \frac{1}{2}n(n-1)\,[\Pr(x_1 - \bar{x} > Y_1)]^2,$ $\Bigg\}$ (5)

so that $\quad \alpha - \frac{1}{2}(n-1)\,\alpha^2/n < \Pr(x_{\max.} - \bar{x} > Y_1) < \alpha.$

We have therefore that for all n and $\alpha = 0{\cdot}05,\ 0{\cdot}01$

$$\left.\begin{matrix} 0{\cdot}04875 \\ 0{\cdot}00995 \end{matrix}\right\} < \Pr(x_{\max.} - \bar{x} > Y_1) < \left\{\begin{matrix} 0{\cdot}05 \\ 0{\cdot}01. \end{matrix}\right.$$

A second approximation $Y_2(\alpha)$, which *underestimates* the true value, is given by solving for Y_2 from

$$T_1(Y_2) = \alpha + T_2(Y_2).$$

This is not very convenient, but we may replace $T_2(Y_2)$ by $\frac{1}{2}(n-1)\,\alpha^2/n$ to obtain a simple and generally very accurate second approximation.

2·1. *The power function of the extreme deviate test*

In using $x_{\max.} - \bar{x}$ to decide whether an outlying observation should be rejected one generally has in mind the following hypotheses:

Null-hypothesis H_0: The $x_t\,(t = 1, \ldots, n)$ are all drawn from a common normal population with unknown mean μ and known variance which may be taken as unity.

* If (4) is applied to the ratio of $x_{\max.} - \bar{x}$ to the standard deviation of the *same* sample of n, then since this ratio is bounded, T_1 will give the left-hand side *exactly* for sufficiently large values of Y (cf. Pearson & Chandra Sekar, 1936).

Alternative hypothesis H_1: One or possibly a few of the x_i come from a normal population with increased mean $\mu+\lambda$ and variance unity, the remaining x_i belonging to the above 'null' population. There could also conceivably be more than one non-null population.

It is possible to apply the test sequentially, that is, having found that the largest observation should be rejected one may proceed to test the second largest, and so on. Such tests have recently been discussed by Hartley (1955),* but we shall confine ourselves to the situation when a single test is carried out. If we take the event A_i to be that $x_i-\bar{x}>Y_\alpha$, then equation (1) will give for $m=1$ the probability of establishing significance. This will now be considered in detail when H_1 specifies only one of the x_i, say x_1, as a true outlier. We have

$$P(\lambda)\equiv \Pr(x_{\max.}-\bar{x}>Y_\alpha \mid H_1)$$
$$=(n-1)\Pr(u_1>Y'_\alpha)+\Pr(u_1>Y'_\alpha-\lambda')-(n-1)\Pr(u_1>Y'_\alpha-\lambda', u_2>Y'_\alpha)-\ldots, \quad (6)$$

where
$$Y'_\alpha=[n/(n-1)]^{\frac{1}{2}}Y_\alpha,$$
$$\lambda'=[n/(n-1)]^{\frac{1}{2}}\lambda,$$

and u_1, u_2 are unit normal variates with correlation $\rho=-1/(n-1)$. From the discussion in the previous section it will be clear that the terms not shown in (6) are negligible if $P(\lambda)$ is required only to moderate accuracy. For $\alpha=0{\cdot}05$ Table 1 shows $P(\lambda)$ for selected values of $n\leqslant 25$ and for $\lambda=1,2,3,4$. The last term of (6) was calculated with the aid of tables given by K. Pearson (1931) and Nicholson (1943).

Table 1. *Probability* $P(\lambda)$ *of rejecting the largest of n observations by the use of* $x_{\max.}-\bar{x}$ *at the 5 % level of significance when all observations are normal with unit variance,* $n-1$ *have mean* μ *and one has mean* $\mu+\lambda$

n \ λ	1	2	3	4
3	0·216	0·654	0·951	0·999
4	0·175	0·557	0·902	0·993
5	0·153	0·496	0·862	0·987
6	0·138	0·453	0·829	0·980
7	0·127	0·420	0·802	0·973
8	0·120	0·395	0·778	0·966
9	0·113	0·374	0·758	0·960
10	0·108	0·357	0·740	0·954
12	0·101	0·329	0·710	0·943
15	0·093	0·300	0·674	0·929
20	0·085	0·266	0·630	0·909
25	0·080	0·244	0·598	0·894

We turn now to the interpretation of $P(\lambda)$. It is a power function as it gives the probability of rejecting H_0 when H_1 is in fact true, and since $P(0)=\alpha$. $P(\lambda)$ is therefore of interest as an indication of the power performance to be expected from the test when the experimenter is merely concerned with receiving a warning about the 'pollution' of his data by at least one

* See also end of §3·1.

rogue observation without wishing to *identify* this rogue. On the other hand, it is *not* equal to the probability of rightly inferring that there *is* a rogue observation *and* identifying it correctly, this being the *joint* probability that the largest observation is a true outlier and that it exceeds Y_α. However, as $\lambda \to 0$ this joint probability tends to approximately α/n and is therefore not a power function, although possessing desirable properties. It is, moreover, difficult to evaluate but will always differ from $P(\lambda)$ by less than α, approaching $P(\lambda)$ (or $\Pr(u_1 > Y'_\alpha - \lambda')$) as λ increases.

3. The maximum F-ratio

Let s_t^2/s^2 $(t = 1 \dots n)$ denote the ratio of two sample variances, which follows an F-distribution with ν_t, ν degrees of freedom. Then the maximum F-ratio is $s^2_{\max.}/s^2$. The importance of this statistic in overcoming the effect of 'selecting' the largest among a number of F-ratios is well known (see, for example, Pearson & Hartley, 1954, p. 39). The cases $\nu_t = 1$, $\nu_t = 2$ have been investigated in detail by Nair (1948*a*) and Finney (1941), respectively, but by different methods from ours. Finney also indicates the solution for other even ν_t. We shall apply equation (3) to the problem.

Take
$$y_t = s_t^2/s^2.$$

Then
$$P_{ij\dots k}(Y) = \Pr\left(\frac{s_i^2}{s^2} > Y, \frac{s_j^2}{s^2} > Y, \dots, \frac{s_k^2}{s^2} > Y\right)$$
$$= \int_0^\infty \Pr(\chi_i^2 > 2Y_i x, \chi_j^2 > 2Y_j x, \dots, \chi_k^2 > 2Y_k x)\, p(x)\, dx,$$

where χ_t^2 is distributed as χ^2 with ν_t degrees of freedom,
$$2Y_t = \nu_t Y/\nu,$$
and
$$p(x) = 2^{-\frac{1}{2}\nu}\Gamma^{-1}(\tfrac{1}{2}\nu)\, x^{\frac{1}{2}\nu - 1} e^{-\frac{1}{2}x} \quad (x \geqslant 0).$$

Writing $Q_t(x) = \Pr(\chi_t^2 > x)$ we have
$$P_{ij\dots k}(Y) = \int_0^\infty \prod_{ij\dots k} Q_t(2Y_t x)\, p(x)\, dx$$
$$= \int_0^\infty \prod_{ij\dots k} e^{-Y_t x}\left[1 + Y_t x + \dots + \frac{(Y_t x)^{\frac{1}{2}\nu_t - 1}}{(\frac{1}{2}\nu_t - 1)!}\right] p(x)\, dx \qquad (7)$$

if the ν_t are even. (7) may be evaluated by termwise integration when the integrand has been multiplied out.

3·1. *Case of equal numerator degrees of freedom*

We consider the following experimental designs of size $l \times l$: (i) the randomized block experiment with equal numbers of treatments and blocks, (ii) the Latin square and (iii) the Graeco-Latin square. Actually, the use of the maximum F-ratio is perhaps even more important in factorial experiments in which a rather larger number of 'treatment' mean squares is tested against a common error mean square. But the designs chosen will serve to illustrate certain interesting points and have the convenience of allowing a fairly systematic tabulation. It should also be noted that in the sequel all numerator mean squares are treated on terms of equality, although it is often more appropriate to apply the method only to those mean squares corresponding to factors of real interest and not to 'extraneous' factors such as block effects.

Table 2 shows the probabilities, obtained from (2) and (7), with which the *ordered* F-ratios $F_{(l)}$ exceed the 5 and 1 % significance levels F_α of the corresponding random F-ratio. As is to be expected, the term $\Pr(F_{(1)} > F_\alpha)$ accounts for by far the largest portion of the 'total Type I error', viz. 2α, 3α, 4α, respectively. This term increases with l for any given design. The ∞-values shown are identical, for any t, with the probabilities obtained by assuming

Table 2. *Showing the probability, for various designs of size $l \times l$, with which the ordered F-ratio $F_{(l)}$ exceeds the upper percentage point F_α of the corresponding random F-ratio*

l	(l)	$\alpha = 0{\cdot}05$			$\alpha = 0{\cdot}01$		
		Randomized block	Latin square	Graeco-Latin square	Randomized block	Latin square	Graeco-Latin square
3	1	0·0842	0·0903	—	0·0172	0·0183	—
	2	0·0158	0·0424	—	0·0028	0·0084	—
	3	—	0·0172	—	—	0·0034	—
5	1	0·0921	0·1238	0·1398	0·0191	0·0265	0·0306
	2	0·0079	0·0230	0·0434	0·0009	0·0031	0·0071
	3	—	0·0032	0·0137	—	0·0003	0·0019
	4	—	—	0·0031	—	—	0·0004
7	1	0·0944	0·1321	0·1623	0·0195	0·0283	0·0359
	2	0·0056	0·0165	0·0315	0·0005	0·0016	0·0036
	3	—	0·0014	0·0055	—	0·0001	0·0004
	4	—	—	0·0006	—	—	0·0000
9	1	0·0953	0·1356	0·1706	0·0197	0·0288	0·0374
	2	0·0047	0·0136	0·0259	0·0003	0·0011	0·0024
	3	—	0·0008	0·0032	—	0·0000	0·0002
	4	—	—	0·0002	—	—	0·0000
∞	1	0·0975	0·1426	0·1855	0·0199	0·0297	0·0394
	2	0·0025	0·0072	0·0140	0·0001	0·0003	0·0006
	3	—	0·0001	0·0005	—	0·0000	0·0000
	4	—	—	0·0000	—	—	0·0000

independence among the variance ratios (Hartley, 1938) and apply to any n-factor design ($n = 2, 3, 4$) with large error degrees of freedom.

In Table 3 we give the upper 5 and 1 % points of the maximum F-ratio in the above cases and also for $l = 6$, 8. Below each exact figure is shown the $100\alpha/n$ % point of the corresponding random F-ratio, this being a simple conservative approximation due to Hartley (1949). The 'exact' values for $l = 6$, 8 were in fact inserted by interpolation with the approximation as an auxiliary function, and may be in error by a few units in the second decimal place. If the error degrees of freedom ν are noted it will be found, as might be expected, that Hartley's approximation increases in accuracy as ν increases (n constant), and as n decreases (ν constant). The usefulness of the approximation, except for very small ν,

is obvious for the cases treated in Table 3. Moreover, in many other situations (including the case of 'extraneous' factors mentioned above) the demands made on the approximation will be less severe.

If the approximation is accepted a very simple procedure permits the complete analysis of the experiment (Hartley, 1955):

Test $F_{(1)}$ by referring to the $100\alpha/n$ level of the corresponding F-ratio. If $F_{(1)}$ is not significant stop, if significant test $F_{(2)}$ at the $100\alpha/(n-1)$ level, etc.

Table 3. *Upper* 100α *percentage points of the maximum F-ratio followed by the corresponding* $100\alpha/n$ *% points of F for the n-factor designs of Table* 2

n / l	$\alpha = 0{\cdot}05$			$\alpha = 0{\cdot}01$		
	2	3	4	2	3	4
3	9·67 10·65	35·4 49·0	— —	24·3 26·3	182·0 299·0	— —
5	3·66 3·73	4·47 4·67	5·86 6·49	5·59 5·64	7·05 7·23	10·2 10·9
6	3·09 3·13	3·55 3·64	(4·09)* (4·31)*	4·40 4·43	5·10 5·17	(6·10)* (6·26)*
7	2·76 2·78	3·08 3·13	3·41 3·51	3·78 3·79	4·21 4·24	4·71 4·76
8	2·18 2·20	2·79 2·82	3·01 3·07	3·39 3·39	3·68 3·70	4·00 4·03
9	2·38 2·40	2·60 2·62	2·76 2·80	3·10 3·10	3·34 3·35	3·56 3·58

* These values must, of course, be interpreted without reference to an experimental design as no 6×6 Graeco-Latin square exists.

3·2. *Generalization to unequal numerator degrees of freedom*

The $F_{\max.}$ test is intuitively at its best when all numerator degrees of freedom ν_t are equal. In the more general case it is still possible to determine exact significance levels of $F_{\max.}$ by means of (7) provided all the ν_t are even. However, the probability with which the tth factor is significant by chance is then no longer independent of t but larger for the smaller ν_t. This difficulty may be overcome by having, for any given design, a specific significance level $F_t(\alpha)$, say, for the tth F-ratio F_t, subject to the following conditions:

(i) probability that F_t exceeds $F_t(\alpha)$ is independent of t;
(ii) probability that at least one F_t exceeds $F_t(\alpha)$ is α.

These conditions result in a set of equations which may be written down by means of (1) and (7) and solved numerically.

Example. In the case of a ν_1+1 by ν_2+1 randomized block experiment $F_1(\alpha)$, $F_2(\alpha)$ are the solutions of

$$\text{(i)} \quad \Pr(s_1^2/s^2 > F_1) = \Pr(s_2^2/s^2 > F_2),$$

$$\text{(ii)} \quad 2\Pr(s_1^2/s^2 > F_1) - \Pr(s_1^2/s^2 > F_1, s_2^2/s^2 > F_2) = \alpha,$$

where s_1^2, s_2^2 and s^2 are mean squares based on ν_1, ν_2 and $\nu_1\nu_2$ degrees of freedom, respectively. For $\nu_1 = 8$, $\nu_2 = 2$ and $\alpha = 0{\cdot}05$ we obtain $F_1(0{\cdot}05) = 3{\cdot}07$, $F_2(0{\cdot}05) = 4{\cdot}58$. These values may be compared with the $2\frac{1}{2}$ % significance levels of the corresponding F-ratios, 3·12, 4·69.

An adequate tabulation would, of course, be feasible only in certain simple cases. But in many practical situations it will be sufficient to test the largest variance ratio against the upper $100\alpha/n$ % point of the appropriate F-ratio. When special accuracy is desired the approximate values will help in the numerical solution for the somewhat smaller exact percentage points. Reference should again be made to Hartley (1955).

I am indebted to the referee for a number of valuable comments on the first draft of this paper and to Miss B. C. Halliburton of the C.S.I.R.O. for her careful computation of the various tables.

REFERENCES

COCHRAN, W. G. (1941). The distribution of the largest of a set of estimated variances as a fraction of their total. *Ann. Eugen., Lond.*, **11**, 47.

FELLER, W. (1950). *An Introduction to Probability Theory and its Applications.* New York: John Wiley and Sons.

FINNEY, D. J. (1941). The joint distribution of variance ratios based on a common error mean square. *Ann. Eugen., Lond.*, **11**, 136.

FISHER, R. A. (1929). Tests of significance in harmonic analysis. *Proc. Roy. Soc.* A, **125**, 54.

GRUBBS, F. E. (1950). Sample criteria for testing outlying observations. *Ann. Math. Statist.* **21**, 27.

HALPERIN, M., GREENHOUSE, S. W., CORNFIELD, J. & ZALOKAR, JULIA (1955). Tables of percentage points for the studentized maximum absolute deviate in normal samples. *J. Amer. Statist. Ass.* **50**, 185.

HARTLEY, H. O. (1938). Studentization and large-sample theory. *J. R. Statist. Soc. Suppl.* **5**, 80.

HARTLEY, H. O. (1949). Tests of significance in harmonic analysis. *Biometrika*, **36**, 194.

HARTLEY, H. O. (1955). Some recent developments in analysis of variance. *Comm. Pure and Appl. Math.* **8**, 47.

MCKAY, A. T. (1935). The distribution of the difference between the extreme observation and the sample mean in samples of n from a normal universe. *Biometrika*, **27**, 466.

NAIR, K. R. (1948*a*). The studentized form of the extreme mean square test in the analysis of variance. *Biometrika*, **35**, 16.

NAIR, K. R. (1948*b*). The distribution of the extreme deviate from the sample mean and its studentized form. *Biometrika*, **35**, 118.

NICHOLSON, C. (1943). The probability integral for two variables, *Biometrika*, **33**, 59.

PEARSON, E. S. & CHANDRA SEKAR, C. (1936). The efficiency of statistical tools and a criterion for the rejection of outlying observations. *Biometrika*, **28**, 308.

PEARSON, E. S. & HARTLEY, H. O. (1954). *Biometrika Tables for Statisticians*, **1**. Cambridge University Press.

PEARSON, K. (1931). *Tables for Statisticians and Biometricians*, **2**. Cambridge University Press.

χ^2 PROBABILITIES FOR LARGE NUMBERS OF DEGREES OF FREEDOM

BY JOHN WISHART

Statistical Laboratory, University of Cambridge

Probability values for χ^2 are given in Pearson & Hartley (1954, Table 7) for χ^2 up to 120 and for degrees of freedom (ν) not greater than 70. This means that the complete range is not covered beyond about $\chi^2 = 40$, and no values are available for really high χ^2. Casting about for a suitable expansion which would make tabulation in this range feasible, the author first had a look at Karl Pearson's modal expansion in terms of incomplete normal moment functions (Pearson, 1922, p. xix), a formula which may alternatively be written in terms of certain χ^2 probabilities, or as a polynomial in a related variate x', combined with chosen values of the normal probability integral and ordinate corresponding to x' ($P(X)$ and $Z(X)$ in the notation of Pearson & Hartley). Karl Pearson calls attention to the slowness of convergence of his formula, and evidently did not recommend it highly for use beyond the tabulated range of the Incomplete Gamma-function Table.

Going back to first principles, we shall find it convenient to adopt the Pearson–Hartley notation of using c for $\frac{1}{2}\nu$. The frequency function of $y = \chi^2/\nu = \frac{1}{2}\chi^2/c$ is

$$\frac{c^c}{\Gamma(c)} y^{c-1} e^{-cy} \quad (0 \leqslant y \leqslant \infty),$$

from which that of $z = \frac{1}{2}\ln y$ is

$$\frac{2c^c}{\Gamma(c)} \exp\{c(2z - e^{2z})\} = \frac{2c^c e^{-c}}{\Gamma(c)} \exp\left[-c\left\{\frac{(2z)^2}{2!} + \frac{(2z)^3}{3!} + \ldots\right\}\right] (-\infty \leqslant z \leqslant \infty). \tag{1}$$

Now write $x = 2z\sqrt{c} = \sqrt{(\frac{1}{2}\nu)}\ln(\chi^2/\nu)$. This leads to the frequency function for x

$$f(x) = \frac{\sqrt{(2\pi)}\, c^{c-\frac{1}{2}} e^{-c}}{\Gamma(c)} \frac{e^{-\frac{1}{2}x^2}}{\sqrt{(2\pi)}} \exp - \left\{\frac{x^3}{3!\, c^{0.5}} + \frac{x^4}{4!\, c} + \frac{x^5}{5!\, c^{1.5}} + \frac{x^6}{6!\, c^2} + \ldots\right\}$$

$$= x_0 \frac{e^{-\frac{1}{2}x^2}}{\sqrt{(2\pi)}} \left\{1 - \frac{x^3}{6c^{0.5}} + \frac{x^6 - 3x^4}{72c} - \frac{5x^9 - 45x^7 + 54x^5}{6{,}480c^{1.5}} + \frac{5x^{12} - 90x^{10} + 351x^8 - 216x^6}{155{,}520c^2} - \ldots\right\}, \tag{2}$$

expanding as far as terms in c^{-2}, to this point the value of x_0 being

$$1 - \frac{1}{12c} + \frac{1}{288c^2}.$$

The range of x is from $-\infty$ to $+\infty$, and $x = 0$, while not the mean of x, corresponds to $\chi^2 = \nu$, i.e. to the mean of χ^2. Effectively we are transforming χ^2 by replacing the ratio of χ^2 to its mean, raised to the power of its standard deviation, by e^{2x}.

To obtain χ^2 probabilities we require to integrate $f(x)\,dx$ from $-\infty$ to X, where X may be positive or negative, or from X to ∞. The answer can be given in more than one form. It may be put in terms of the incomplete normal moment functions

$$\mu_n(X) = \frac{1}{\sqrt{(2\pi)}} \int_0^X x^n e^{-\frac{1}{2}x^2} dx,$$

$$m_n(X) = \frac{\mu_n(X)}{(n-1)(n-3)\ldots 1 \text{ or } 2},$$

according as n is even or odd, for comparison with Karl Pearson's modal expansion. We then get

Formula A. $X = \sqrt{c}\ln(\chi^2/(2c))$:

$$\int_0^X f(x)\,dx = \left(1-\frac{1}{12c}+\frac{1}{288c^2}\right)\left[\mu_0(X)-\frac{m_3(X)}{3c^{0\cdot5}}+\frac{1}{c}\left\{\tfrac{5}{24}m_6(X)-\tfrac{1}{8}m_4(X)\right\}\right.$$
$$-\frac{1}{c^{1\cdot5}}\left\{\tfrac{8}{27}m_9(X)-\tfrac{1}{3}m_7(X)+\tfrac{1}{15}m_5(X)\right\}$$
$$\left.+\frac{1}{c^2}\left\{\tfrac{385}{1152}m_{12}(X)-\tfrac{35}{64}m_{10}(X)+\tfrac{91}{384}m_8(X)-\tfrac{1}{48}m_6(X)\right\}\right] \quad (3)$$

as far as terms in c^{-2}. Values of m_1 to m_{12} are given to seven decimals in the old Tables (Pearson, 1914, 1931), while $\mu_0(X) = P(X) - 0\cdot5$ and $P(X)$ is readily available, either in the old Tables or in the new *Biometrika Tables*.

From this formula we can derive a very useful expression for the chance of χ^2 being less than, or at most equal to, its degrees of freedom ν. For this will be $\int_{-\infty}^{0} f(x)\,dx$, obtained from Formula A by putting $x = -\infty$, and noting that

$$m_{2r+1}(-\infty) = -(2\pi)^{-\frac{1}{2}} \quad \text{and} \quad m_{2r}(-\infty) = 0\cdot5.$$

Formula A'.

$$\int_{-\infty}^{0} f(x)\,dx = \left(1-\frac{1}{12c}+\frac{1}{288c^2}\right)\left[\frac{1}{2}+\frac{1}{\sqrt{(2\pi)}}\frac{1}{3c^{0\cdot5}}+\frac{1}{24c}+\frac{1}{\sqrt{(2\pi)}}\frac{4}{135c^{1\cdot5}}+\frac{1}{576c^2}\right]$$
$$= \frac{1}{2}+\frac{1}{3\sqrt{(2\pi c)}}\left(1+\frac{1}{180c}\right). \quad (4)^*$$

This formula will give seven-decimal accuracy for c of the order of 50 or above, i.e. beyond the range of the Incomplete Gamma-function Tables (the p of that table is our $c-1$). Adding or subtracting (3) and (4) gives the required χ^2 probabilities.

Alternatively, we may write in (3)

$$\sqrt{(2\pi)}\,m_{2r+1}(X) = P(X^2 \mid 2r+2),$$
$$2m_{2r}(X) = P(X^2 \mid 2r+1),$$

where $P(X^2 \mid \nu)$ denotes the probability that χ^2 does not exceed X^2, for ν degrees of freedom. These probabilities may be obtained to five decimals by subtracting from unity the χ^2 probabilities given in the *Biometrika Tables*, Table 7.

By integrating the terms of (2) by parts from $-\infty$ to X, we may get the probability that x does not exceed X in the form of an expansion in powers of X, involving

$$P(X) = \frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{X} e^{-\frac{1}{2}x^2}\,dx$$

and

$$Z(X) = \frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}X^2}.$$

Multiplying in by the outside factor we get our result in the form of

Formula B. $X = \sqrt{c}\ln(\chi^2/(2c))$:

$$\int_{-\infty}^{X} f(x)\,dx = P(X)+Z(X)\left\{\frac{X^2+2}{6c^{0\cdot5}}-\frac{X^5+2X^3+6X}{72c}+\frac{5X^8-5X^6+24X^4+6X^2+12}{6{,}480c^{1\cdot5}}\right.$$
$$\left.-\frac{5X^{11}-35X^9+36X^7-144X^5-180X^3-540X}{155{,}520c^2}\right\}. \quad (5)$$

* An additional term (inside the bracket) is $-25/(2{,}016c^2)$.

Now the well-known Cornish–Fisher expansion (Cornish & Fisher, 1937) applies to the general z function. We may adapt this to our special case by writing $n_1 = \nu = 2c$, $n_2 = \infty$. In the formulae for $a, b, \ldots, f$ on p. 30·13 of the reference, we have $\delta = \sigma = (2c)^{-1}$, and if we then substitute in the formulae on p. 30·9 we get *Formula B* above, Cornish & Fisher's p, z and ξ being replaced by $P(X)$, $Z(X)$ and X respectively. Thus the expansion we have obtained is not a new one, but is a special case of a *direct* Cornish–Fisher expansion in which n_2 is allowed to go to infinity. Our derivation of this expansion from the frequency function of χ^2 by a few lines of algebra is, however, believed to be new.

The best-known form of the Cornish–Fisher expansion is the *inverse* one, namely, a formula which determines z_0 in terms of X, where the chance that z is less than, or at most equal to, z_0 is equal to $P(X)$. A chosen value of $P(X)$ determines X from normal probability tables. We can therefore get the inverse form to *Formula B* by putting $\delta = \sigma = (2c)^{-1}$ in the formulae on p. 30·14 of the reference cited, but note that in the penultimate line δ^4 should read δ^2. With $z = \frac{1}{2}\ln(\chi^2/\nu)$ we get

Formula B′.

$$z_0 = \frac{X}{2c^{0\cdot5}} - \frac{X^2+2}{12c} + \frac{X^3+5X}{72c^{1\cdot5}} - \frac{6X^4+59X^2+58}{3{,}240c^2} + \frac{9X^5+232X^3+599X}{77{,}760c^{2\cdot5}}. \qquad (6)$$

Corresponding to a chosen value of the probability that χ^2 be not exceeded, we then obtain the corresponding χ^2 as νe^{2z_0} from tables of natural logarithms.

A footnote to Table 8 in the *Biometrika Tables* (Pearson & Hartley, 1954) gives two approximations for $\nu > 100$. One which is practically as good as the Wilson–Hilferty cube-root approximation, and better than Fisher's $\sqrt{(2\chi^2)}$, can be obtained from the first two terms of *Formula B′* by putting $\ln(\chi^2/\nu) = X\sqrt{(2/\nu)} - (X^2+2)/(3\nu)$. Equally, a fair approximation to the probabilities beyond the range of Table 7 may be got from the first two terms of (5).

Examples.

(1) $\nu = 98$, $c = 49$. Let $\chi^2 = 98$. Use *Formula A′*.

$$P(\chi^2 \leqslant 98) = \frac{1}{2} + \frac{1}{21\sqrt{(2\pi)}}\left(1 + \frac{1}{8820}\right)$$
$$= 0{\cdot}5189\,994.$$

From the Incomplete Gamma-function Tables

$$u = \chi^2/\sqrt{(2\nu)} = 7, \quad p = 48, \quad I(u,p) = 0{\cdot}5189\,993.$$

(2) $\nu = 100$, $c = 50$. Let $\chi^2 = 124{\cdot}342$. Use *Formula B*.

$$z = \tfrac{1}{2}\ln 1{\cdot}24342 = 0{\cdot}1089\,33, \quad X = 2z\sqrt{c} = 1{\cdot}5405\,453,$$

$$\begin{aligned} P(\chi^2 \leqslant 124{\cdot}342) &= 0{\cdot}9382\,8625 \\ &+ \quad 125\,5249 \\ &- \quad 8\,5353 \\ &+ \quad 1346 \\ &+ \quad 98 \\ &\quad 0{\cdot}9499\,996 \end{aligned}$$

Biometrika Tables (Table 8) give 0·05 point for χ^2 as 124·342.

(3) $\nu = 100$, $c = 50$. Let $Q(\chi_0^2) = P(\chi^2 \geqslant \chi_0^2) = 0{\cdot}05$. Use *Formula B'*.

$$\begin{aligned} X &= 1{\cdot}6448\,5363 \\ z_0 &= 0{\cdot}1163\,0872 \\ &- \quad 784257 \\ &+ \quad 49790 \\ &- \quad 3229 \\ &+ \quad 155 \\ &\underline{\quad 0{\cdot}1089\,333} \end{aligned}$$ giving $\chi_0^2 = 124{\cdot}3421$.

(4) $\nu = 100$, $c = 50$. Let $\chi^2 = 77{\cdot}9295$. Use *Formula B*.

$$z = -0{\cdot}1246\,828, \quad X = -1{\cdot}7632\,811,$$

$$\begin{aligned} P(\chi^2 \leqslant 77{\cdot}9295) &= 0{\cdot}0389\,2655 \\ &+ \quad 101\,5025 \\ &+ \quad 90351 \\ &+ \quad 2132 \\ &- \quad 123 \\ &\underline{\quad 0{\cdot}0500\,004} \end{aligned}$$

Biometrika Tables (Table 8) give 0·95 point for χ^2 as 77·9295.

(5) $\nu = 100$, $c = 50$. Let $P(\chi_0^2) = P(\chi^2 \leqslant \chi_0^2) = 0{\cdot}05$. Use *Formula B'*.

$$\begin{aligned} z_0 &= -0{\cdot}1163\,0872 \\ &- \quad 784257 \\ &- \quad 49790 \\ &- \quad 3229 \\ &- \quad 155 \\ &\underline{-0{\cdot}1246\,830} \end{aligned}$$ giving $\chi_0^2 = 77{\cdot}9295$.

The urge to undertake this investigation might have been lacking had Slutskii's Tables (1950) been available earlier to the author. For Slutskii tabulates the χ^2 probabilities corresponding to $t = \sqrt{(2\chi^2)} - \sqrt{(2\nu)}$, first for $\nu = 6$ to 32 (Table III) and secondly for $\nu(2/\nu) = 0$ to 0·25, i.e. for $\nu = \infty$ down to 32 (Table IV). The tables are given to five decimal places, and Table IV, in particular, is a very useful supplement to Pearson & Hartley's Table 7. Formulae A and B above may, nevertheless, be useful for tabulating χ^2/ν probabilities to considerable accuracy, or, in abbreviated form, for obtaining approximate probabilities without the use of tables. They are also interesting in their own right. They are special cases of a general expansion which the author has since worked out, and which will be given in a separate paper.

REFERENCES

CORNISH, E. A. & FISHER, R. A. (1937). *Rev. Inst. int. Statist.* **4**, 307. (*Contributions to Mathematical Statistics*, by R. A. Fisher, paper 30. 1950. London: Chapman and Hall.)

PEARSON, E. S. & HARTLEY, H. O. (1954). *Biometrika Tables for Statisticians*, **1**. Cambridge University Press.

PEARSON, K. (1914, 1931). *Tables for Statisticians and Biometricians*. Part I (Table IX), Part II (Table XIII). London: Biometrika Office.

PEARSON, K. (1922). *Tables of the Incomplete* Γ*-function*. Reissued 1934, 1946. Cambridge University Press.

SLUTSKII, E. E. (1950). *Tables for the Calculation of the Incomplete* Γ*-function and the* χ^2 *Probability Function*. Moscow and Leningrad: Isdatelstvo Akademii Nauk SSSR.

THE SAMPLING DISTRIBUTION OF A MAXIMUM-LIKELIHOOD ESTIMATE

By J. B. S. HALDANE and SHEILA MAYNARD SMITH
University College London

The method of maximum likelihood was used by Edgeworth (1908) and others. Since Fisher's (1921) paper, which gave a general expression for the variance of an estimate by it, it has been very widely used. However, no serious attempt to determine the form of its sampling distribution when the number in the sample is not very large was made in the next thirty years, though Hotelling (1930) and others improved the rigour of earlier proofs. Haldane (1953*a*) gave its bias, but gave an incorrect expression for its third moment. These were, however, obtained as special cases of a more general theorem. The paper is, moreover, hard to follow owing to numerous misprints, and several errors, some, at least, of which are here corrected. In this paper we obtain the first four moments of the distribution by a more direct approach.

Throughout we consider the estimation of a single unknown parameter ξ. Let x be the maximum-likelihood estimate of ξ, based on a sample of N members. Before considering samples from a continuous distribution, we suppose that the population sampled can be classified into a finite or denumerable set of classes, that the number of individuals in the sample falling into the rth class is n_r, and that the expectation of n_r on some hypothesis is

$$\mathscr{E}(n_r) = Nf_r(\xi).$$

The permissible range of values of ξ is that for which no $f_r(\xi)$ is negative. We assume that over this range, which may consist of several discrete parts, every $f_r(\xi)$ is regular. In particular, we assume that, even at the boundaries of this range, ξ remains finite.

If this last restriction is not made, some or all the moments of the sampling distribution of x are infinite. Thus if

$$\mathscr{E}(n_1) = \frac{N\xi}{1+\xi}, \quad \mathscr{E}(n_2) = \frac{N}{1+\xi}, \quad \text{then} \quad x = \frac{n_1}{n_2}.$$

Hence $\mathscr{E}(x)$ is infinite, since n_2 has a finite probability of being zero. In such estimations the maximum-likelihood estimate is seriously biased even if cases where one class is empty are excluded. Thus in the above example x still has a bias $+N^{-1}\xi(1+\xi)+O(N^{-2})$ when cases with $n_2 = 0$ are omitted, whilst $x' = n_1/(n_2+1)$ has a bias tending to zero more rapidly than any negative power of N. This criticism applies to some other efficient estimators, for example, the 'product method' which Fisher (1954) recommends for linkage estimation.

If

$$f_r(\xi) = \binom{m}{r} \xi^r (1-\xi)^{m-r},$$

then $x = \sum_r rn_r/(mN)$, and its sampling distribution is binomial, with cumulants

$$\kappa_1 = \xi, \quad \kappa_2 = N^{-1}\xi(1-\xi), \quad \kappa_3 = N^{-2}\xi(1-\xi)(1-2\xi), \text{etc.}$$

In this case, the maximum-likelihood estimate is absolutely unbiased. The same is true for a Poisson distribution, or for a distribution in which $h+k\xi$ is substituted for ξ in $f_r(\xi)$ above. We conjecture that it is biased in all other cases except for special values of ξ. It will be

shown that in the important case where expectations are linear functions of ξ, that is to say, $f_r(\xi) = h_r + k_r\xi$, the bias is of order N^{-2}, and thus usually negligible, whereas in other cases it tends to zero with N^{-1}.

SYMBOLISM AND PRELIMINARY LEMMATA

Let $$f_r(\xi) = a_r, \quad f_r'(\xi) = b_r, \quad f_r''(\xi) = c_r, \quad f_r'''(\xi) = d_r, \quad f_r^{iv}(\xi) = e_r.$$

Since $\sum_r f_r(u) \equiv 1$ for all u, it follows that $\sum_r a_r = 1$, $\sum_r b_r = \sum_r c_r$ etc. $= 0$. Clearly a_r, b_r, etc., are constants over any series of samples from a given population, though their values may have to be estimated.

Let $x = \xi + y$, so that y is the error of the estimate.

Let $n_r = N(a_r + z_r)$.

Let $$A_i = \sum_r a_r^{-i} b_r^{i+1}, \quad B_i = \sum_r a_r^{-i} b_r^i c_r, \quad C_i = \sum_r a_r^{-i} b_r^{i-1} c_r^2,$$

$$D_i = \sum_r a_r^{-i} b_r^i d_r, \quad S_i = \sum_r h_r^i a_r, \text{ where } h_r \text{ is arbitrary.}$$

Let $$\alpha_i = \sum_r a_r^{-i} b_r^{-i} z_r, \quad \beta_i = \sum_r a_r^{-i} b_r^{i-1} c_r z_r, \quad \delta_i = \sum_r a_r^{-i} b_r^{i-1} d_r z_r,$$

$$\theta = \alpha_2 - \beta_1, \quad \lambda = 2\alpha_3 - 3\beta_2 + \delta_1.$$

Suffixes will be dropped when the meaning is clear. Thus $\Sigma a^{-1} bz$ means $\sum_r a_r^{-1} b_r z_r$, and so on.

It is at once obvious from Taylor's theorem that

$$\left.\begin{aligned} f_r(x) &= a_r + b_r y + \tfrac{1}{2} c_r y^2 + \tfrac{1}{6} d_r y^3 + \ldots, \\ f_r'(x) &= b_r + c_r y + \tfrac{1}{2} d_r y^2 + \tfrac{1}{6} e_r y^3 + \ldots. \end{aligned}\right\} \tag{1}$$

The symbolism is that of Haldane (1953*a*). The quantities A_i, etc., are connected by simple relationships such as

$$\frac{d}{d\xi} A_i = -iA_{i+1} + (i+1) B_i, \quad \frac{d}{d\xi} B_i = -iB_{i+1} + iC_i + D_i.$$

A_i and C_i are positive when i is odd. A_i becomes large when any a_r is small, or any b_r large, that is to say, for values of ξ for which any $f_r(\xi)$ is small, or is changing rapidly. B_i and D_i are usually large under the same conditions, and when a first or second derivative of any $f_r(\xi)$ is changing rapidly. Such a term as $A_1^{-\frac{3}{2}} A_2$, which occurs in the expression for the skewness of the distribution of x, is usually large if any $f_r(\xi)$ is small, and so on.

In order to obtain expressions for such expectations as $\mathscr{E}(\alpha_1^4)$ and $\mathscr{E}(\alpha_1^5\theta)$ we require general expressions for the expectations of powers and products such as $(\Sigma h_r z_r)^4$ and $(\Sigma h_r z_r)^5 \Sigma k_r z_r$, where h_r and k_r are constants. First consider the sampling distribution of $\sum_r h_r z_r$. Then if $S_i = \sum_r h_r^i a_r$ it is easy to show, from the moments of the multinomial distribution, that the expectation of $\sum_r h_r z_r$ is zero, and its first few moments are:

$$\left.\begin{aligned} \mu_2 &= N^{-1}(S_2 - S_1^2), \\ \mu_3 &= N^{-2}(S_3 - 3S_1S_2 + 2S_1^3), \\ \mu_4 &= 3N^{-2}(S_2 - S_1^2)^2 + N^{-3}(S_4 - 4S_1S_3 - 3S_2^2 + 12S_1^2S_2 - 6S_1^4), \\ \mu_5 &= 10N^{-3}(S_2 - S_1^2)(S_3 - 3S_1S_2 + 2S_1^3) + O(N^{-4}), \\ \mu_6 &= 15N^{-3}(S_2 - S_1^2)^3 + O(N^{-4}). \end{aligned}\right\} \tag{2}$$

For example,

$$\begin{aligned}
\mathscr{E}(z_r^3) &= N^{-2}a_r(1-a_r)(1-2a_r),\\
\mathscr{E}(z_r^2 z_s) &= -N^{-2}a_r a_s(1-2a_r),\\
\mathscr{E}(z_r z_s z_t) &= 2N^{-2}a_r a_s a_t.\\
(\Sigma h_r z_r)^3 &= \Sigma h_r^3 z_r^3 + 3\Sigma h_r^2 h_s z_r^2 z_s + 6\Sigma h_r h_s h_t z_r z_s z_t.
\end{aligned}$$

So
$$\begin{aligned}
\mathscr{E}[(\Sigma h_r z_r)^3] &= N^{-2}[\Sigma h_r^3(a_r - 3a_r^2 + 2a_r^3) + 3\Sigma h_r^2 h_s(-a_r a_s + 2a_r^2 a_s) + 12\Sigma h_r h_s h_t a_r a_s a_t]\\
&= N^{-2}[\Sigma h_r^3 a_r - 3\Sigma h_r a_r \Sigma h_r^2 a_r + 2(\Sigma h_r a_r)^3]\\
&= N^{-2}(S_3 - 3S_1 S_2 + 2S_1^3).
\end{aligned}$$

The leading terms of μ_4, μ_5 and μ_6 can be derived directly from μ_2 and μ_3. From (2) we can at once obtain the expectations of powers. For

$$\alpha_1 = \sum_r a_r^{-1} b_r z_r, \quad \text{so} \quad h_r = a_r^{-1} b_r,$$

and
$$S_1 = 0, \quad S_2 = \Sigma a_r^{-1} b_r^2 = A_1, \quad S_3 = A_2, \quad S_4 = A_3.$$

So from (2),
$$\mathscr{E}(\alpha_1^4) = 3N^{-2}A_1^2 + N^{-3}(A_3 - 3A_1^2).$$

To obtain expectations of products, the expressions (2) must be expanded if necessary. For example

$$\mu_5 = 2N^{-3}(5S_2 S_3 - 5S_1^2 S_3 - 15S_1 S_2^2 + 25S_1^3 S_2 - 10S_1^5) + O(N^{-4}).$$

In $\mathscr{E}[(\Sigma h_r z_r)^4 (\Sigma k_r z_r)]$ one of the five h's in turn is replaced by k_r. Thus

$$\begin{aligned}
\mathscr{E}[(\Sigma h_r z_r)^4 (\Sigma k_r z_r)] = 2N^{-3}[&3(S_2 - S_1^2)(\Sigma h^2 k a - S_2 \Sigma k a)\\
&+ 2(S_3 - 6S_1 S_2 + 5S_1^3)(\Sigma h k a - S_1 \Sigma k a)] + O(N^{-4}).
\end{aligned}$$

From such expressions we immediately evaluate expectations of products. For example, in

$$\alpha_1^4\theta = \alpha_1^4(\alpha_2 - \beta_1), \quad h = a^{-1}b, \quad k = a^{-2}b^2 - a^{-1}c.$$

So
$$S_1 = 0, \quad S_2 = A_1, \quad S_3 = A_2, \quad \Sigma h k a = A_2 - B_1, \quad \Sigma h^2 k a = A_3 - B_2, \quad \Sigma k a = A_1.$$

The required expectations are as follows:

$$\left.\begin{aligned}
\mathscr{E}(\alpha_1) &= 0,\\
\mathscr{E}(\alpha_1^2) &= N^{-1}A_1, \quad \mathscr{E}(\alpha_1\theta) = N^{-1}(A_2 - B_1),\\
\mathscr{E}(\alpha_1^3) &= N^{-2}A_2, \quad \mathscr{E}(\alpha_1^2\theta) = N^{-2}(A_3 - B_2 - A_1^2),\\
\mathscr{E}(\alpha_1^4) &= 3N^{-2}A_1^2 + N^{-3}(A_3 - 3A_1^2),\\
\mathscr{E}(\alpha_1^3\theta) &= 3N^{-2}A_1(A_2 - B_1) + O(N^{-3}),\\
\mathscr{E}(\alpha_1^3\lambda) &= 3N^{-2}A_1(2A_3 - 3B_2 + D_1) + O(N^{-3}),\\
\mathscr{E}(\alpha_1^2\theta^2) &= N^{-2}[2(A_2 - B_1)^2 + A_1(A_3 - 2B_2 + C_1) - A_1^3] + O(N^{-3}),\\
\mathscr{E}(\alpha_1^5) &= 10N^{-3}A_1 A_2 + O(N^{-4}),\\
\mathscr{E}(\alpha_1^4\theta) &= 2N^{-3}[2A_2(A_2 - B_1) + 3A_1(A_3 - B_2) - 3A_1^3] + O(N^{-4}),\\
\mathscr{E}(\alpha_1^6) &= 15N^{-3}A_1^3 + O(N^{-4}),\\
\mathscr{E}(\alpha_1^5\theta) &= 15N^{-3}A_1^2(A_2 - B_1) + O(N^{-4}),\\
\mathscr{E}(\alpha_1^5\lambda) &= 15N^{-3}A_1^2(2A_3 - 3B_2 + D_1) + O(N^{-4}),\\
\mathscr{E}(\alpha_1^4\theta^2) &= 3N^{-3}A_1[4(A_2 - B_1)^2 + A_1(A_3 - 2B_2 + C_1) - A_1^3] + O(N^{-4}),
\end{aligned}\right\} \quad (3)$$

Haldane (1953*a*) gave an incorrect expression for $\mathscr{E}[(\Sigma hz)^2(\Sigma kz)^2]$ in his equation (4). His equation (6) should read

$$\mathscr{E}(\alpha_1^2\eta_1) = N^{-2}(2A_2 + A_1H_1).$$

CALCULATION OF THE MOMENTS AND CUMULANTS

The maximum-likelihood estimate is a root of

$$\sum_r n_r f_r'(x)\,[f_r(x)]^{-1} = 0. \tag{4}$$

There is rarely if ever any doubt as to which root is the estimator. Substituting from (1), provided that a_r is not zero, and summing, we obtain

$$\alpha_1 - (A_1+\theta)y + \tfrac{1}{2}(2A_2 - 3B_1 + \lambda)y^2 - \tfrac{1}{6}(6A_3 - 12B_2 + 3C_1 + 4D_1)y^3 + \ldots = 0.$$

These terms are sufficient to calculate the fourth moment of y to order N^{-3}. The first term tends to zero with $N^{-\frac{1}{2}}$, while the coefficients of powers of y tend to constant values with large N. So when N is large we may invert this series, which is the expansion of

$$\sum_r \left[a_r^{-1}\frac{d}{dy}\ln f_r(\xi+y)\right],$$

by Lagrange's theorem.

We thus find

$$\begin{aligned} y = {} & A_1^{-1}\alpha_1 + A_1^{-3}[(A_2 - \tfrac{3}{2}B_1)\alpha_1 - A_1\theta]\alpha_1 \\ & + A_1^{-5}[\{2(A_2 - \tfrac{3}{2}B_1)^2 - A_1(A_3 - 2B_2 + \tfrac{1}{2}C_1 + \tfrac{2}{3}D_1)\}\alpha_1^2 \\ & - 3A_1(A_2 - \tfrac{3}{2}B_1)\alpha_1\theta + \tfrac{1}{2}A_1^2\alpha_1\lambda + A_1^2\theta^2]\alpha_1 + O(N^{-3}). \end{aligned} \tag{5}$$

From this fundamental equation and equations (3) the moments of the distribution of y are calculated. In order to evaluate κ_4 we need two terms in the expansion of μ_2.

$$\mathscr{E}(y) = -\tfrac{1}{2}N^{-1}A_1^{-2}B_1 + O(N^{-2}).$$

This is the bias of x:

$$\mathscr{E}(y^2) = N^{-1}A_1^{-1} + N^{-2}A_1^{-4}[-A_2^2 + \tfrac{15}{4}B_1^2 + A_1(A_3 - B_2 - D_1) - A_1^3] + O(N^{-3}).$$

Subtracting $[\mathscr{E}(y)]^2$,

$$\mu_2 = N^{-1}A_1^{-1} + N^{-2}A_1^{-4}[-A_2^2 + \tfrac{7}{2}B_1^2 + A_1(A_3 - B_2 - D_1) - A_1^3] + O(N^{-3}).$$

Thus A_1 is the amount of information about ξ per individual in the sample:

$$\mathscr{E}(y^3) = N^{-2}A_1^{-3}(A_2 - \tfrac{9}{2}B_1) + O(N^{-3}).$$

Subtracting $3\mathscr{E}(y)\mathscr{E}(y^2) - 2[\mathscr{E}(y)]^3$,

$$\mu_3 = N^{-2}A_1^{-3}(A_2 - 3B_1) + O(N^{-3}),$$

$$\mathscr{E}(y^4) = 3N^{-2}A_1^{-2} + N^{-3}A_1^{-5}[-6A_2^2 - 14A_2B_1 + \tfrac{105}{2}B_1^2 + A_1(7A_3 - 6B_2 - 10D_1) - 9A_1^3] + O(N^{-4}).$$

Subtracting $4\mathscr{E}(y)\mathscr{E}(y^3) - 6\mathscr{E}(y^2)[\mathscr{E}(y)]^2 + 3[\mathscr{E}(y)]^4$,

$$\mu_4 = 3N^{-2}A_1^{-2} + N^{-3}A_1^{-5}[-6A_2^2 - 12A_2B_1 + 45B_1^2 + A_1(7A_3 - 6B_2 - 10D_1) - 9A_1^3] + O(N^{-4}),$$

$$\begin{aligned} \kappa_4 &= \mu_4 - 3\mu_2^2 \\ &= N^{-3}A_1^{-5}[-12B_1(A_2 - 2B_1) + A_1(A_3 - 4D_1) - 3A_1^3] + O(N^{-4}). \end{aligned}$$

Thus the first four cumulants of the distribution of x are:

$$\left.\begin{aligned}
\kappa_1 &= \xi - \tfrac{1}{2}N^{-1}A_1^{-2}B_1 + O(N^{-2}),\\
\kappa_2 &= N^{-1}A_1^{-1} + N^{-2}A_1^{-4}[-A_2^2 + \tfrac{7}{2}B_1^2 + A_1(A_3 - B_2 - D_1) - A_1^3] + O(N^{-3}),\\
\kappa_3 &= N^{-2}A_1^{-3}(A_2 - 3B_1) + O(N^{-3}),\\
\kappa_4 &= N^{-3}A_1^{-5}[-12B_1(A_2 - 2B_1) + A_1(A_3 - 4D_1) - 3A_1^3] + O(N^{-4}),
\end{aligned}\right\} \tag{6}$$

whence

$$\begin{aligned}
\gamma_1 &= N^{-\frac{1}{2}}A_1^{-\frac{3}{2}}(A_2 - 3B_1) + O(N^{-\frac{3}{2}}),\\
\gamma_2 &= N^{-1}A_1^{-3}[-12B_1(A_2 - 2B_1) + A_1(A_3 - 4D_1) - 3A_1^3] + O(N^{-2}).
\end{aligned}$$

If every $f_r(\xi)$ is linear, that is to say, of the form $k_r + b_r\xi$, then every c_r and d_r, and hence every B_i, C_i and D_i vanishes, and the maximum-likelihood estimate is almost unbiased. It has, however, a bias of order N^{-2}, evaluated by Haldane (1953*a*) and given below. This can perhaps always be neglected. The cumulants of the distribution of x are:

$$\left.\begin{aligned}
\kappa_1 &= \xi + 2N^{-2}A_1^{-5}(A_2^3 - 2A_1A_2A_3 + A_1^2A_4) + O(N^{-3}),\\
\kappa_2 &= N^{-1}A_1^{-1} - N^{-2}A_1^{-4}(A_1^3 - A_1A_3 + A_2^2) + O(N^{-3}),\\
\kappa_3 &= N^{-2}A_1^{-3}A_2 + O(N^{-3}),\\
\kappa_4 &= N^{-3}A_1^{-4}(A_3 - 3A_1^2) + O(N^{-4}),
\end{aligned}\right\} \tag{7}$$

whence

$$\begin{aligned}
\gamma_1 &= N^{-\frac{1}{2}}A_1^{-\frac{3}{2}}A_2 + O(N^{-\frac{3}{2}}),\\
\gamma_2 &= N^{-1}(A_1^{-2}A_3 - 3) + O(N^{-2}).
\end{aligned}$$

Thus γ_2 is positive near the limits of the range of ξ, where at least one value of a_r is small, and hence $A_1^{-2}A_3$ is large. It may be negative in the middle of the range.

Now in practice the constants A_i, B_i, etc., are unknown. They must be estimated from x, the estimate of ξ. If we put A_1' for the maximum-likelihood estimate of A_1, namely,

$$\sum_r [f_r'(x)]^2 [f_r(x)]^{-1},$$

then

$$A_1' = A_1 - (A_2 - 2B_1)\,y + (A_3 - \tfrac{5}{2}B_2 + C_1 + D_1)\,y^2 + \ldots.$$

So

$$\mathscr{E}(A_1') = A_1 + N^{-1}A_1^{-1}[A_1^{-1}B_1(\tfrac{1}{2}A_2 - B_1) + A_3 - \tfrac{5}{2}B_2 + C_1 + D_1] + O(N^{-2}),$$

and similar expressions can be found for other estimates. Thus (6) and (7) remain true when A_i is substituted for A_i and so on, except that the second term in the sampling variance has no validity.

We may also use the fact that $f_r(x) = N^{-1}n_r + O(N^{-1})$ and use the approximations

$$\begin{aligned}
A_i'' &= N^i \Sigma n_r^{-i}[f_r'(x)]^{i+1}, & B_i'' &= N^i \Sigma n_r^{-i}[f_r'(x)]^i f_r''(x),\\
C_i'' &= N^i \Sigma n_r^{-i}[f_r'(x)]^{i-1}[f_r''(x)]^2, & D_i'' &= N^i \Sigma n_r^{-i}[f_r'(x)]^i f_r'''(x)
\end{aligned}$$

for A_i, B_i, C_i and D_i in expressions (6) and (7) without introducing serious errors provided no n_r is small.

We have so far neglected the possibility that any a_r is zero. This can only be the case if, in equation (4), at least one n_r is zero, though it may be necessary that several n_r's should vanish. Thus if

$$f_1(x) = x, \quad f_2(x) = x^2, \quad f_3(x) = 1 - x - x^2,$$

equation (4) gives

$$(n_1 + 2n_2 + 2n_3)\,x^2 + (n_1 + 2n_2 + n_3)\,x - (n_1 + 2n_2) = 0,$$

and since no n_r can be negative, x can only be zero if $n_1 = n_2 = 0$.

If some a_r, say a_1, is zero, then n_1 is zero. If no other a_r is zero, (4) becomes $f_1(x) = 0$, and the relevant solution is $x = \xi$, where $f_1(\xi) = a_1 = 0$. That is to say, the estimate is absolutely unbiased. Since $a_1 = 0$, A_1 is infinite, that is to say, x is not only an unbiased estimate, but a precise estimate. However, we can never know that this is the case on the basis of a finite sample. The numbers A_i, B_i, etc., all become infinite if any a_r vanishes, and the estimates of them become meaningless if any $f_r(x)$ is zero, and unreliable if it is sufficiently small. This is of course paralleled in the case of simple sampling when no members of a sample possess some attribute. This problem has been discussed since the time of Laplace, and a further discussion would be out of place here. To sum up, if there is a set of one or more values of n_r whose expectation is zero for some value of ξ, then if all these values are zero the expressions found are inapplicable, and if their total is small (say less than 5) they become very inaccurate.

So far we have supposed that the sample is a sample of a discontinuous variate. Let us now consider a continuous variate p, whose distribution is given by:

$$dF = f(p, \xi)\, dp,$$

where f is a known function, and ξ an unknown parameter to be estimated. Suppose that in a sample of N members p has assumed the values $p_1, p_2, \ldots, p_r, \ldots, p_N$, then the maximum-likelihood estimator is

$$\sum_{r=1}^{N} f'(p_r, x)\,[f(p_r, x)]^{-1} = 0,$$

and we have only to put

$$a'_r = f(p_r, x), \quad b'_r = \frac{\partial}{\partial x} f(p_r, x), \quad c'_r = \frac{\partial^2}{\partial x^2} f(p_r, x), \quad d'_r = \frac{\partial^3}{\partial x^3} f(p_r, x),$$

and $A_i = \Sigma a'^{-i}_r b'^{i+1}_r$, etc., to be able to use formulae (6) and (7). It follows that the maximum-likelihood estimate of the mean of a normal distribution is absolutely unbiased, and that if $f(p, \xi)$ is of the form

$$\xi f_1(p) + (1 - \xi) f_2(p),$$

where $f_1(p)$ and $f_2(p)$ are known frequency functions, the maximum-likelihood estimate of ξ has a bias tending to zero with N^{-2}. In practice, unless N were quite small, we should group our data, and $f_r(x)$ would be $f(p_r, x)$, where p_r is the central value of p in the rth interval.

DISCUSSION

We first correct an error of Haldane (1953*a*). In his equations (12) the expression for the third cumulant of the generalized estimate should read

$$\kappa_3 = A_1^{-3} N^{-2} [A_2 - 3B_1] + O(N^{-3}).$$

Since κ_3 is independent of k the skewness of the distribution of the minimum discrepancy estimate ($k = +1$), or of the estimate found by putting $k = 2$ so that

$$\sum_r [n_r(n_r + 1) f'_r(x) \{f_r(x)\}^{-2}] = 0,$$

which is nearly equivalent to minimizing χ^2, is the same as the skewness of the maximum-likelihood estimate.

The statistical importance of a knowledge of the distribution of x can best be judged from examples. Let us consider the case where

$$f_1(x) = \tfrac{1}{4}(2 + x), \quad f_2(x) = \tfrac{1}{2}(1 - x), \quad f_3(x) = \tfrac{1}{4}x,$$

which arises in linkage estimation from F_2 data, and has been extensively discussed by Fisher (1954) and Haldane (1953*a*). The maximum-likelihood solution is the positive root of $Nx^2-(n_1-2n_2-n_3)x-2n_3=0$, an alternative efficient estimate by the method of minimum discrepancy (Haldane, 1953*a*) being

$$\frac{2(n_3+1)(2n_1-n_2+1)}{(n_1+1)(n_2+1)+4(n_1+1)(n_3+1)+(n_2+1)(n_3+1)}.$$

Here B_i, C_i, etc., are zero and we can use (7):

$$A_1=\tfrac{1}{2}(1+2x)[x(1-x)(2+x)]^{-1},\quad A_2=\tfrac{1}{2}(2-2x-5x^2-4x^3)[x(1-x)(2+x)]^{-2},$$
$$A_3=\tfrac{1}{2}(4-6x-3x^2+14x^3+12x^4+6x^5)[x(1-x)(2+x)]^{-3}.$$

Hence the bias of x is $72x(2+x)(1-x)(1+2x)^{-5}N^{-2}$, which becomes $+144xN^{-2}$ when x is very small, and $+\frac{8}{9}(1-x)N^{-2}$ when x is nearly unity. It is always negligible.

The variance, as is well known, is $2x(1-x)(2+x)[(1+2x)N]^{-1}$ and

$$\gamma_1=(2-2x-5x^2-4x^3)[\tfrac{1}{2}x(1-x)(1+2x)^3(2+x)N]^{-\frac{1}{2}},$$
$$\gamma_2=(8-18x-27x^2+19x^3+48x^4+24x^5)[x(1-x)(2+x)(1+2x)^2]^{-1}N^{-1}.$$

Thus γ_1 changes sign when $x=0{\cdot}418$. γ_2 is negative if x lies between 0·353 and 0·611. Thus over this range the maximum-likelihood estimate is rather more precise than it would be were it normally distributed. When x or $1-x$ is small, it is less precise. When $x=\frac{1}{2}$ and $N=100$, $\gamma_1=0{\cdot}0474$ and $\gamma_2=-0{\cdot}0065$.

As a non-linear example consider the case where, of N organisms subject to constant mortality, n_1 have died during the first of three equal intervals, n_2 during the second, and n_3 during the third, while n_4 survive. If ξ be the probability of surviving through an interval,

$$f_1(x)=1-x,\quad f_2(x)=x(1-x),\quad f_3(x)=x^2(1-x),\quad f_4(x)=x^3.$$

Thus
$$x=\frac{n_2+2n_3+3n_4}{n_1+2n_2+3n_3+3n_4},$$
$$A_1=(1+x+x^2)[x(1-x)]^{-1},\quad A_2=(1+2x+2x^2-8x^3)[x(1-x)]^{-2},$$
$$A_3=(1+8x+9x^2-58x^3+43x^4)[x(1-x)]^{-3},\quad B_1=2(1+2x)[x(1-x)]^{-1},$$
$$B_2=2(3+7x-13x^2)[x(1-x)]^{-2},\quad D_1=6[x(1-x)]^{-1}.$$

Thus the bias of x is $-x(1-x)(1+2x)(1+x+x^2)^{-2}N^{-1}$, which is always negative, and the variance $x(1-x)(1+x+x^2)^{-1}N^{-1}$. Thus if $x=\frac{1}{2}$ and $N=100$, the bias is $-0{\cdot}00163$ and the standard error 0·0378. It is barely worth while to correct for a bias of 4 % of the standard error.

$$\gamma_1=[Nx(1-x^3)]^{-\frac{1}{2}}(1+x+x^2)^{-1}(1-4x-4x^2+4x^3),$$

which changes sign when $x=0{\cdot}214$, and is only $-0{\cdot}13$ when $x=\frac{1}{2}$ and $N=100$.

$$\gamma_2=N^{-1}(1+x+x^2)^{-3}[x(1-x)]^{-1}$$
$$\times[1-18x+12x^2+118x^3-57x^4-84x^5+28x^6+6x^7+3x^8]+O(N^{-2}).$$

γ_2 is negative if x lies between 0·059 and 0·351. Over the rest of the range it is positive, and the maximum-likelihood estimate is somewhat less precise than a normally distributed one. When x is small γ_2 approximates to $(Nx)^{-1}$, and when $1-x$ is small, to $[3N(1-x)]^{-1}$. When $x=\frac{1}{2}$, $\gamma_2=1039/(343N)+O(N^{-2})$. Thus when $x=\frac{1}{2}$ and $N=100$, $\gamma_2=+0{\cdot}0303$.

Haldane (1953*b*) considered the estimation by maximum likelihood of several unknown parameters, and obtained an expression for the bias when two were estimated, as well as

confirming the known expressions for their variances and covariance. The method used was the same as that employed in this paper. The calculation of the higher cumulants of the joint distribution of estimates is clearly, however, very complicated.

Fisher (1921) showed that if ϕ is any regular monotonic function, and x is the maximum-likelihood estimate of ξ, $\phi(x)$ is the maximum-likelihood estimate of $\phi(\xi)$. We can now ask the further question: For what functions is $\phi(x)$ an almost unbiased estimate of $\phi(\xi)$? Haldane (1953*a*) defined an almost unbiased estimate as one whose bias tends to zero with N^{-2} or more rapidly as N increases:

$$\begin{aligned}\phi(x) &= \phi(\xi+y)\\ &= \phi(\xi)+y\phi'(\xi)+\tfrac{1}{2}y^2\phi''(\xi)+\ldots.\end{aligned}$$

So
$$\begin{aligned}\mathscr{E}[\phi(x)] &= \phi(\xi)+\phi'(\xi)\,\mathscr{E}(y)+\tfrac{1}{2}\phi''(\xi)\,\mathscr{E}(y^2)+\ldots\\ &= \phi(\xi)-\tfrac{1}{2}N^{-1}A_1^{-2}B_1\phi'(\xi)+\tfrac{1}{2}N^{-1}A_1^{-1}\phi''(\xi)+O(N^{-2}).\end{aligned}$$

So the condition is

$$\frac{\phi''(\xi)}{\phi'(\xi)} = A_1^{-1}B_1.$$

Hence
$$\phi(\xi) = \int \exp\left[\int A_1^{-1}B_1\,d\xi\right] d\xi.$$

For example, in the linear case here considered x is an almost unbiased estimate of ξ. In the non-linear case

$$A_1^{-1}B_1 = \frac{2(1+2\xi)}{1+\xi+\xi^2}, \quad \text{so} \quad \phi(\xi) = E+F\xi(1+\xi+\xi^2+\tfrac{1}{2}\xi^3+\tfrac{1}{5}\xi^4),$$

where E and F are any constants.

We shall briefly discuss the value of unbiased estimates in the branches of biology with which we are familiar. If the estimate, so to speak, stands by itself, a maximum-likelihood estimate is probably as useful as an unbiased one. The same is true when a function of the estimate will later be tabulated. If, on the other hand, the estimate is to be one of a large collection, say gene frequencies in populations, or relative viabilities of phenotypes, it is undesirable that the estimates should be consistently high or low, on the average, since a mean or median based on a large number of such estimates will have a bias which in some cases could be of the order of magnitude of its own standard error. It is therefore desirable that estimates which are subsequently to be tabulated should be unbiased.

SUMMARY

Expressions are given for the first four cumulants of a maximum-likelihood estimate. The bias is only negligible in some such estimates, when the sample number is of the order of 100.

REFERENCES

EDGEWORTH, F. V. (1908). On the probable error of frequency constants. *J. R. Statist. Soc.* **71**, 381.
FISHER, R. A. (1921). On the mathematical foundations of theoretical statistics. *Phil. Trans.* A, **222**, 308.
FISHER, R. A. (1954). *Statistical Methods for Research Workers*, 12th ed. London: Oliver and Boyd.
HALDANE, J. B. S. (1953*a*). A class of efficient estimates of a parameter. *Bull. Int. Stat. Inst.* **33**, 231.
HALDANE, J. B. S. (1953*b*). The estimation of two parameters from a sample. *Sankhyā*, **12**, 313.
HOTELLING, H. (1930). The consistency and ultimate distribution of optimum statistics. *Trans. Amer Math. Soc.* **32**, 847.

TESTS FOR RANDOMNESS OF POINTS ON A LINE

By D. E. BARTON and F. N. DAVID
University College London

1. In previous papers (1956*a*, *b*) we have discussed the distribution of various functions of ordered intervals any of which might be used as a test for the randomness of points on a line. In this present note we proceed a stage further and discuss approximate distributions of sums of ordered and unordered intervals. These distributions can be used to test the randomness of points on a line or as bivariate tests for randomness. The power function of any, or all, of these tests with respect to a specified alternative can be obtained to a similar degree of approximation and we propose to describe these in a further note.

2. It is possible to rank intervals along a line according to their position from one end or according to their magnitude. We assume that on a line of unit length, $(n-1)$ points are dropped randomly giving n random intervals. Let the points be dropped at distances $\{x_i\}$ from one end $(i = 1, 2, \ldots (n-1))$, with

$$x_1 \leqslant x_2 \leqslant \ldots \leqslant x_{n-1}.$$

Write $$d_1 = x_1, \quad d_i = x_i - x_{i-1} \quad (1 < i \leqslant (n-1)), \quad d_n = 1 - x_{n-1},$$

so that d_i is the ith interval along the line. These $\{d_i\}$ can be given a rank according to their magnitude. Letting the smallest have rank 1 and the largest rank n we may write these as $\{g_i\}$. The positional rank of the interval g_i which is the ith largest in size we write as r_i. The series $\{g_i\}$ which is merely a rearrangement of the series $\{d_i\}$ will be

$$g_1 \leqslant g_2 \leqslant \ldots \leqslant g_n.$$

The magnitudinal rank of d_i which we do not use except for computational rearrangement we may write as r'_i. We have then that

$$d_i = g_{r_i} \quad \text{and} \quad g_i = d_{r'_i}.$$

To fix ideas we have given a numerical example in Table 1. (This example is not ideal in that there are several intervals of the same size. Further investigation into the dates of the consecrations would have determined the exact magnitude of the intervals, but since we are illustrating only, we assigned intervals of the same size a random order.) Using the position and size of the ith interval in position and the position and size of the ith interval in magnitude, four criteria suggest themselves:

$$\text{(i)} \quad R = \sum_{i=1}^{n} i r'_i.$$

$$\text{(ii)} \quad Y = \sum_{i=1}^{n} i d_i.$$

$$\text{(iii)} \quad G = \sum_{i=s}^{n} r_i g_i.$$

$$\text{(iv)} \quad G^* = n \sum_{i=1}^{n} d_i g_i.$$

3. R as a test for randomness has been discussed on several occasions, the first, we think, being by Hotelling & Pabst (1936) in a bivariate application. They showed that R is approximately distributed as a Pearson Type I. Olds (1938) tabulated a related quantity for values of n from 2 to 10 after which the result that $Rn^{\frac{1}{2}}$ is normally distributed can be used. The exact distribution of Y is immediate. We have

$$Y = \sum_{i=1}^{n} id_i = n - \sum_{i=1}^{n-1} x_i.$$

The x_i are $(n-1)$ independent rectangular variables, and since we consider them all the fact that they are ordered will play no part in the distribution of their sum. We have, using the result of Hall (1927) and Irwin (1927), that

$$p(Y) = \frac{1}{(n-1)!} \sum_{j=0}^{k} (-1)^{j+1\,n-1}C_j[n-j-Y]^{n-2}$$
$$[(n-k) > Y > (n-k-1);\ k = 1, 2, \ldots, (n-1)].$$

The exact distribution is not, however, of great importance, since the mean of $(n-1)$ independent rectangular variables tends very quickly to be normally distributed with increasing n. We have approximately that

$$Y' = 1 - \frac{Y}{n}$$

is a normal variable with

$$\mathscr{E}(Y') = \frac{n-1}{2n}, \quad \text{var}\,(Y') = \frac{n-1}{12n^2},$$

which is accurate enough for $n \geqslant 25$. For n small we may use the fact that the percentage points for a rectangular sum are embodied in Barton's (1953) tables of ψ_1^2. For

$$\frac{12}{n-1}\left(\frac{n+1}{2} - Y\right)^2$$

is exactly distributed as ψ_1^2. If a one-tailed test is required we may use the fact that Y has a symmetrical distribution. The criterion Y will obviously be sensitive to departures from randomness of the type where the probability is changing along the line; in fact, as we shall discuss in a further paper, under certain circumstances it will be the most powerful test which can be used. It will not be sensitive to fluctuations of the type where a long interval tends to be followed by a short one.

4. If we allow s to equal unity then G is Y. It occurred to us that the discussion of a partial sum of intervals, such as that defined by G, might be useful in that the zero intervals arising from lack of refined measurement could thereby be eliminated. (This lack of refined measurement will lead to a bias in that it may overemphasize the larger intervals, but on the whole this bias does not seem to be of serious proportions.) The derivation of the distribution of G is theoretically possible following the algebra developed in our two earlier papers (1956a, b). Since the derivation is the same, consider a more general form of G say

$$G_{st} = \sum_{i=s}^{t} r_i g_i \quad (s \leqslant t \leqslant n).$$

Under the null hypothesis r_i and g_i are independent and we have

$$p(r_s r_{s+1} \ldots r_t g_s g_{s+1} \ldots g_t) = p(r_s r_{s+1} \ldots r_t)\, p(g_s g_{s+1} \ldots g_t)$$
$$\propto p(r_s r_{s+1} \ldots r_t) \sum_{j=0}^{s-1} (-1)^{j\,s-1}C_j(1-(j+1)g_s - g_{s+1} - \ldots - g_{t-1} - (n-t+1)g_t)^{n-t+s-2}.$$

Substitute for g_t:

$$p(r_s r_{s+1} \dots r_t g_s g_{s+1} \dots g_{t-1} G_{st}) \propto p(r_s r_{s+1} \dots r_t) \sum_{j=0}^{s-1} \frac{(-1)^j}{r_t}\, {}_{s-1}C_j \left[1 - \frac{(n-t+1)}{r_t} G_{st} + g_s\left(\frac{(n-t+1)r_s}{r_t} - (j+1)\right) + \dots + g_{t-1}\left(\frac{(n-t+1)}{r_t} r_{t-1} - 1\right)\right]^{n-t+s-2}.$$

Integrate out for $g_{t-1}, \dots, g_s$ in turn and at any stage the inequalities

$$\frac{G_{st} - r_s g_s - r_{s+1} g_{s+1} - \dots - r_{k-1} g_{k-1}}{r_t + r_{t-1} + \dots + r_k} > g_k > g_{k-1}, \quad \frac{G_{st}}{\sum_{l=s}^{t} r_l} > g_s > 0 \quad (k = t-1, t-2, \dots, s+1)$$

will hold. The conditional distribution for G_{st} given a set of ranks is thus proportional to

$$\sum_{j=1}^{s-1} (-1)^j\, {}_{s-1}C_j \sum_{w=1}^{t-s} \frac{(-1)^{w-1}\left(\sum_{f=t-w+1}^{t} r_f\right)^{t-s-1} \quad A}{\prod_{l=1}^{t-s-w}\left[\sum_{i=t-l-w+1}^{t-w} r_i(n-t+w) - l \sum_{i=t-w+1}^{t} r_i\right] \prod_{v=1}^{w-1}\left[\sum_{i=t-l-w+1}^{t-v} (n-t+v) r_i - (w-v) \sum_{i=t-v+1}^{t} r_i\right]} \times \left[(n-t+w) r_s - (t-s-w+j+1) \sum_{f=t-w+1}^{t} r_f\right],$$

where

$$A = \left(1 - \frac{(n-s+j+1)}{\sum_{f=s}^{t} r_f} G_{st}\right)^{n-2} - \left(1 - \frac{(n-t+w)}{\sum_{f=t-w+1}^{n-t+w} r_f} G_{st}\right)^{n-2}.$$

The complete distribution of G_{st} will be obtained by summing this conditional distribution over all possible permutations of the ranks $\{r_i\}$. Since we have been unable to simplify the conditional distribution we have not pursued the complete distribution as it will obviously be intractable. Previous work (Barton & David, 1956*b*) on the unweighted partial sums of the g's showed that the normal approximation was fairly good for $n \geqslant 20$, and it might be anticipated that the same would be true for this present case. We show in § 7 for certain special cases that the β_1 and $\beta_2 - 3$ are of order (n^{-1}).

5. The moments of G_{st} may be derived from those we have already given (1956*a*) for the powers and products of $\{g_i\}$ together with the moments of the finite population of natural numbers given in Table 2. Let

$$M(g_r^h) = \mathscr{E}(g_r - \mathscr{E}(g_r))^h$$

and $\kappa(g_r^h)$ the corresponding cumulant. We have

$$\mathscr{E}(G_{st}) = \tfrac{1}{2}(n+1) \sum_{i=s}^{t} \kappa(g_i),$$

where

$$\kappa(g_i) = \frac{1}{n} \sum_{j=1}^{i} \frac{1}{n-j+1},$$

$$\mathscr{E}(G_{st}^2) = \frac{2n+1}{6n}\left[\sum_{s}^{t} S(i^2) + \sum_{s}^{t} (S(i))^2\right] + \frac{3n+2}{6n}\left[\sum_{s}^{t} S(i^2)(s-i) + \frac{1}{2}\left[\sum_{s}^{t} S(i)\right]^2 - \frac{1}{2}\sum_{s}^{t} (S(i))^2\right],$$

where

$$S(i^a) = \sum_{j=1}^{i} \frac{1}{(n-j+1)^a}.$$

In terms of the cumulants of the individual g's this second crude moment may be written

$$\mathscr{E}(G_{st}^2) = \frac{(n+1)(2n+1)}{6}\sum_{s}^{t}\kappa(g_i^2) + \frac{(n+1)(3n+2)}{6}\sum_{i<k}\kappa(g_i g_k) + \frac{(n+1)(3n+2)}{12}[\sum_{i}\kappa(g_i)]^2 + \frac{n(n+1)}{12}\sum_{i}[\kappa(g_i)]^2.$$

Values of the expected value of G, i.e. of G_{st} when $t = n$, and of its standard error for different values of n and s are given in Table 3 below. They are multiplied by n^{-1} to facilitate comparison with their asymptotic form.

6. The higher moments of G_{st} can also be obtained by the same method. Thus if

$$S^{(q)}_{m_1 m_2 \dots m_n} = \sum_{i=s}^{t} S(i^{m_1}) \dots S(i^{m_n})(t-i)\dots(t-i-q+1),$$

we have

$$\mathscr{E}(G_{rs}^3) = \frac{(n+1)}{4(n+2)}[2S_3 + S_{21} + 2S_2S_1 + \tfrac{1}{2}S_1^3 + \tfrac{1}{2}S_{11}S_1 + 8S_3^{(1)} + 2S_2^{(1)} + 3S_1S_2^{(1)} + S_2^{(1)} + 6S_3^{(2)} + (n-t)(S_{21} - S(t)S_2)],$$

with a more complicated expression for the fourth crude moment which we do not reproduce here. In the case in which we are particularly interested, that is, when t equals n, the moments may be considerably simplified by expressing them as polynomials in $S((s-1)^u)$, with coefficients as rational functions of n and $k = n-s+1$. The $S((s-1)^u)$ are just the differences of the appropriate polygamma function of n and k. Thus if

$$1 + S((s-1)^1) = 1 + F(n) - F(k) = T_1 \quad \text{(say)},$$

$$S((s-1)^2) = -F(n) + F(k) = T_2 \quad \text{(say)},$$

and generally

$$S((s-1)^u) = \frac{(-1)^{u-1}}{(u-1)!}\left[\frac{d^u}{dx^u}\log\Gamma(x+1)\right]_k^n = T_u,$$

the central moments of G are (or follow from)

$$\mu_1' = \frac{n+1}{2n}kT_1,$$

$$\mu_2 = \frac{k}{12n^2}[T_1^2(n^2-4nk-3k) + nT_2(3nk+2k+n) + 2n(2n+1)],$$

$$\mu_3 = \frac{(n+1)k}{4n^3(n+2)}[T_3n^3k(k+1) + T_1T_2nk(n^2-n-4nk-2k) + nT_1(n^2-5nk-2k) + T_1^3k(3nk+2k-n^2) + 2n^3],$$

$$\mu_4' = \frac{k}{240n(n+2)(n+3)}\{4(88n^3+138n^2+23n-12) + 4k(20n^3+24n^2-5n-6) + 16T_1[n(9n^2+15n+4) + k(15n^3+21n-4)] + 4(T_1^2+T_2)[2n^2(n+2) + nk(40n^2+69n+23) + k^2(30n^3+35n^2-11n-12)] + (T_1^4+6T_1^2T_2+3T_2^2+8T_1T_3+6T_4)[-2n^2(n+1) + nk(5n^2+9n+2) + 2nk^2(15n^2+25n+8) + k^3(15n^3+15n^2-10n-8)]\}.$$

7. We now consider the case where the ratio s/n is fixed and equal to p say. Let

$$q = 1-p, \quad Q = 1-\log q.$$

Since the Euler–Maclaurin expansion of T_u yields

$$T_1 = Q+O\left(\frac{1}{n}\right), \quad T_{u+1} = \frac{u!}{n^u q^u}(1-q^u)+O\left(\frac{1}{n^{u+1}}\right) \quad (u \geqslant 1),$$

we have for the cumulants of G

$$\mu_1' = \frac{nq}{2}Q+O(1), \quad \kappa_2 = \frac{nq}{12}(7-3q+Q^2(1+4q))+O(1),$$

$$\kappa_3 = \frac{nq}{8}[5-q^2+4(1-5q+2q^2)Q+2q(3q-1)Q^3]+O(1),$$

$$\kappa_4 = \frac{nq}{240}[(537-1464q+1140q^2-245q^3)+12Q(17-92q-5q^2+20q^3)$$
$$+4Q^2(5-271q+914q^2-350q^3)-2Q^4(1+4q-168q^2+384q^3)]+O(1).$$

These reduce to the values for the rectangular mean when $q = 1 = Q$. A comparison of (i) the actual values of $n\beta_1$ and $n(\beta_2-3)$ calculated from the formulae of § 6 with (ii) the first term of their expansion as a power series in n^{-1} using the formulae immediately above is given in Table 3.

8. It is possible to obtain the moments of G using a moment-generating function of a certain simplicity. From our previous work, if W_{st} denotes the weighted sum of $g_s, \ldots, g_t$ with weights $r_1, \ldots, r_k$, we have

$$\sum_{m=0}^{\infty} \mathscr{E}(W_{st}^m)^{n+m-1}C_m y^m = \mathscr{E}\{(1-W_{st}y)^{-n}\}$$
$$= \left[\left(1-\frac{r_k y}{n-t+1}\right)\left(1-\frac{(r_k+r_{k-1})}{n-t+2}y\right)\ldots\left(1-y\frac{\sum_1^k r_i}{n-s+1}\right)\ldots\left(1-y\frac{\sum_1^k r_i}{n}\right)\right]^{-1}.$$

Hence if the $\{r_i\}$ are randomly drawn from the finite population of the first n natural numbers independently of the $\{g_i\}$, if $t = n$, and

$$\bar{r}_j = \frac{1}{j}(r_1+\ldots+r_j),$$

we have

$$\mathscr{E}(G^m) = \frac{m!\,(n-1)!}{(n+m-1)!}\sum_{v=0}^{m} k^v h_v\left(\frac{1}{n-i+1}\right)\mathscr{E}\{h_{m-v}(\bar{r}_i)\,\bar{r}_k^v\},$$

where

$$h_v(\alpha_i) = \sum_{i_1 \leqslant i_2 \leqslant \ldots \leqslant i_v} \alpha_{i_1}\ldots\alpha_{i_v}$$

is the homogeneous product sum of weight v. Let

$$E_{ij\ldots k} = \mathscr{E}(\bar{r}_i\bar{r}_j\ldots\bar{r}_k) \quad (i \leqslant j \leqslant \ldots \leqslant k).$$

We can work out simply the corresponding h-functions which are seen to reduce to linear sums of the moments of the r's whose coefficients are polynomials in k. The E's are given as

functions of i, j, l. and the moments of the finite population in Table 4, as they are quite general and of some intrinsic interest. The quantities

$$h_v\left(\frac{1}{n-i+1}\right)$$

are simply expressed in terms of the corresponding power sums (David & Kendall, V, 1955) which are just the $\{T_u\}$ and the moments of the finite population of the first n natural numbers (Table 2). The moments of G follow immediately after a little reduction.

The asymptotic values for the mean and variance agree well enough for $n = 25$ for them to be used for $n > 25$ in forming the test function in place of the true moments based on the polygammas. Of course if tables of these last are easily available they may be used.

9. The sum of squares of intervals along a line has been discussed by various authors, the first being Greenwood (1946). Approximations to its distribution have been suggested, but the exact distribution has not yet been obtained except in the cases $n = 2, 3$ (Moran, 1947). It was therefore to be expected that the distribution of G^* would be equally elusive, and this has in fact proved to be the case. The exact distributions of G^* for $n = 2, 3$ have been found, but they proved to be of much greater complexity than those found by Moran for the sum of squares. The moments, however, are not difficult. Straightforward algebra gives

$$\mathscr{E}(G^*) = 1, \quad \text{var}\,(G^*) = (n^2+7n-6)/(n+1)\,(n+2)\,(n+3).$$

The shape constants are

$$\beta_1 = \frac{16(n+1)\,(n+2)\,(n+3)\,(n^3+51n^2-133n+60)^2}{(n-4)^2\,(n+5)^2\,(n^2+7n-6)^3} \qquad (n \geqslant 3)$$

$$= 0 \quad (n = 2),$$

$$\beta_2 = \frac{3(n+1)\,(n+2)\,(n+3)\,(n^5+56n^4+4639n^3-15364n^2+17172n-5040)}{(n+4)\,(n+5)\,(n+6)\,(n+7)\,(n^2+7n-6)^2} \qquad (n \geqslant 4)$$

$$= \tfrac{67}{14} \quad (n = 3)$$

$$= \tfrac{25}{9} \quad (n = 2).$$

β_1 for $n = 2$ and β_2 for $n = 2, 3$ were derived separately from the distributions and are not obtainable from the general formula. β_1 and β_2 for various values of n were calculated. They are found to behave in similar fashion to those of Greenwood's sum of squares; as n increases they diverge from the normal point and then turn back approaching normality very slowly. Table 5 gives a selection of these values. Approximations to the percentage points of these distributions can be made by using the appropriate Pearson curves with the first four correct moments or Johnson's system. The percentage points actually quoted in Table 6 were obtained by interpolation into Table 42 of Pearson & Hartley (1954), and outside the range of the tables were estimated from a Johnson S_U curve. The standardized percentage points are also given to make interpolation easier for intermediate values of n.

10. Earlier we remarked that the criterion Y could be expected to be sensitive to departures from randomness where the probability is changing monotonically along the unit line. This remark will also hold good for G. G^*, on the other hand, like Greenwood's sum of squares, may be expected to be sensitive to any type of departure from randomness in which the intervals become either too regular or too irregular. For interest, we have computed the four criteria which we mentioned in § 2 for the data in Table 1.

We have

(i) $R = 87770$. Standardized $R = 2{\cdot}1350$. Standardized R is approximately a unit normal variable, and this value of R is therefore probably significant.

(ii) $Y = 38{\cdot}0693$, $Y' = 0{\cdot}440$, Mean $Y' = 0{\cdot}493$, $\sigma_{Y'} = 0{\cdot}03475$ which makes the hypothesis of randomness doubtful but not impossible, the equivalent normal deviate being $-1{\cdot}52$.

(iii) For G we decided, for illustrative purposes, to cut off observations of magnitudinal rank 13 and under, thus omitting one-fifth, approximately, of the observations.

$G = 37{\cdot}657$ and for 68 observations *in toto* it appears legitimate to use the asymptotic mean and variance. Thus, approximately

Mean $G = 33{\cdot}272$, $\sigma_G = 2{\cdot}436$, giving an equivalent normal deviate of $1{\cdot}817$ which is possibly significant.

(iv) Finally, $G^* = 1{\cdot}119$ which is inside the upper 5 % level of significance as Table 6 shows.

All four criteria, although clearly not independent of each other, point in the same direction, that of possibly rejecting the hypothesis of randomness and of assuming that the probability may be changing with time. Other and simpler tests which could be devised will, no doubt, lead to the same conclusion. Which, among the many possible tests, one would use in practice can only be determined after a discussion of the power with respect to defined alternatives.

Table 1. *Dates of consecration of Archbishops of Canterbury from Thomas à Becket*

Year	X_i	d_i	r_i / i	r'_i	Year	X_i	d_i	r_i / i	r'_i	Year	X_i	d_i	r_i / i	r'_i
1162	0				1366	204				1716	554			
		12	1	37			2	26	8			21	51	58
1174	12				1368	206				1737	575			
		11	2	34			7	27	27			10	52	30
1185	23				1375	213				1747	585			
		6	3	24			6	28	25			10	53	33
1191	29				1381	219				1757	595			
		2	4	13			16	29	48			1	54	6
1193	31				1397	235				1758	596			
		12	5	36			17	30	51			10	55	32
1205	43				1414	252				1768	606			
		2	6	12			29	31	67			15	56	47
1207	45				1443	281				1783	621			
		21	7	57			9	32	29			22	57	59
1228	66				1452	290				1805	643			
		1	8	7			2	33	9			23	58	61
1229	67				1454	292				1828	666			
		2	9	18			32	34	68			20	59	54
1231	69				1486	324				1848	686			
		3	10	15			25	35	64			14	60	40
1234	72				1511	349				1862	700			
		11	11	35			2	36	10			6	61	22
1245	83				1513	351				1868	706			
		25	12	65			20	37	53			15	62	44
1270	108				1533	371				1883	721			
		3	13	16			23	38	62			13	63	38
1273	111				1556	394				1896	734			
		5	14	20			3	39	17			7	64	26
1278	116				1559	397				1903	741			
		1	15	4			16	40	49			25	65	63
1279	117				1575	413				1928	766			
		15	16	43			8	41	28			14	66	41
1294	132				1583	421				1942	780			
		19	17	52			21	42	56			3	67	14
1313	151				1604	442				1945	783			
		0	18	1			6	43	23			10	68	31
1313	151				1610	448				1955	793			
		15	19	46			23	44	60					
1328	166				1633	471								
		5	20	21			27	45	66					
1333	171				1660	498								
		15	21	45			3	46	18					
1348	186				1663	501								
		1	22	5			15	47	42					
1349	187				1678	516								
		0	23	2			13	48	39					
1349	187				1691	529								
		17	24	50			4	49	19					
1366	204				1695	533								
		0	25	3			21	50	55					

Note. $X_i = 793x_i$ where x_i is defined in §2.

Table 2. *The crude moments of the finite population of natural numbers* $1, 2, \ldots, n$

$$\mu'_1 = \frac{n+1}{2}, \qquad \mu'_2 = \frac{(n+1)(2n+1)}{6}, \qquad \mu'_{11} = \frac{(n+1)(3n+2)}{12},$$

$$\mu'_3 = \frac{n(n+1)^2}{4}, \qquad \mu'_{21} = \frac{n(n+1)^2}{6}, \qquad \mu'_{111} = \frac{n(n+1)^2}{8},$$

$$\mu'_4 = \frac{(n+1)}{30}(6n^3+9n^2+n-1), \quad \mu'_{31} = \frac{(n+1)}{120}(15n^3+21n^2-4), \quad \mu'_{22} = \frac{(n+1)}{180}(20n^3+24n^2-5n-6),$$

$$\mu'_{211} = \frac{(n+1)}{360}(30n^3+35n^2-11n-12), \qquad \mu'_{1111} = \frac{(n+1)}{240}(15n^3+15n^2-10n-8).$$

Table 3. *Values of* μ'_1/n, μ_2/n, $n\beta_1$ *and* $n(\beta_2-3)$ *of the distribution of* G *for different* n *and* p $(= s/n)$ *together with their limiting forms*

	p \ n	5	10	15	20	25	∞
μ'_1/n	$\frac{1}{5}$	0·6000	0·5445	0·5261	0·5168	0·5113	0·4893
	$\frac{2}{5}$	0·5760	0·5144	0·4940	0·4838	0·4776	0·4532
	$\frac{3}{5}$	0·5220	0·4525	0·4294	0·4179	0·4109	0·3833
	$\frac{4}{5}$	0·4280	0·3458	0·3179	0·3038	0·2953	0·2609
μ_2/n	$\frac{1}{5}$	0·06667	0·07960	0·08033	0·08205	0·08309	0·08724
	$\frac{2}{5}$	0·07493	0·08752	0·09172	0·09383	0·09510	0·10022
	$\frac{3}{5}$	0·09238	0·10626	0·11082	0·11309	0·11445	0·11989
	$\frac{4}{5}$	0·11316	0·12315	0·12570	0·12680	0·12740	0·12936
$n\beta_1$	$\frac{1}{5}$	0·000	0·007	0·013	0·016	0·018	0·029
	$\frac{2}{5}$	0·069	0·141	0·174	0·192	0·204	0·259
	$\frac{3}{5}$	0·289	0·480	0·571	0·624	0·659	0·824
	$\frac{4}{5}$	0·734	1·293	1·587	1·772	1·846	2·516
$n(\beta_2-3)$	$\frac{1}{5}$	−1·500	−1·733	−1·772	−1·814	−1·839	−1·975
	$\frac{2}{5}$	−2·470	−2·708	−2·809	−2·868	−2·958	−3·087
	$\frac{3}{5}$	−2·826	−2·809	−2·816	−2·814	−2·814	−2·815
	$\frac{4}{5}$	−2·082	−1·506	−1·184	−0·978	−0·836	−0·288

Table 4. *The crude moments of members of a sequence of partial means (population finite)*

$$E_i = \mu'_1,$$

$$E_{ij} = \frac{1}{j}[\mu'_2 + (j-1)\,\mu'_{11}] \quad (i \leqslant j),$$

$$E_{ijk} = \frac{1}{jk}[\mu'_3 + (2(j-1)+(k-1))\,\mu'_{21} + (j-1)(k-2)\,\mu'_{111}] \quad (i \leqslant j \leqslant k),$$

$$E_{ijkl} = \frac{1}{jkl}[\mu'_4 + (2(j-1)+(k-1)+(l-1))\,\mu'_{31} + (2(j-1)+(k-1))\,\mu'_{22}$$
$$+ (3(j-1)(k-2) + 2(j-1)(l-2) + (k-1)(l-2))\,\mu'_{211} + (j-1)(k-2)(l-3)\,\mu'_{1111}] \quad (i \leqslant j \leqslant k \leqslant l).$$

Table 5. *β_1, β_2 for the distribution of G^*, sample size increasing*

n	2	3	4	5	6	7	8	9	10	20	40	100	400	1000	∞
β_1	0	0·96	1·97	2·66	3·08	3·30	3·38	3·36	3·29	2·06	0·90	0·26	0·05	0·02	0
β_2	2·78	4·79	7·17	9·16	10·66	11·74	12·46	12·92	13·17	11·54	7·67	4·51	3·26	3·09	3

Table 6. *Percentage points of G^**

[$\overline{G}^*$ = standardized G^*.]

Percentage point	$n = 40$		$n = 100$		$n = 400$		$n = 1000$	
	G^*	$\overline{G}^*$	G^*	$\overline{G}^*$	G^*	$\overline{G}^*$	G^*	$\overline{G}^*$
Upper 0·5	1·591	3·71	1·325⁻	3·24	1·142	2·84	1·086	2·71
1	1·488	3·06	1·278	2·77	1·126	2·51	1·077	2·43
2·5	1·362	2·27	1·217	2·17	1·103	2·07	1·064	2·02
5	1·273	1·72	1·172	1·72	1·085⁻	1·69	1·053	1·68
Lower 5	0·776	−1·41	0·849	−1·51	0·921	−1·58	0·949	−1·61
2·5	0·726	−1·72	0·819	−1·80	0·906	−1·87	0·940	−1·90
1	0·657	−2·16	0·782	−2·17	0·889	−2·21	0·929	−2·25
0·5	0·602	−2·50	0·755⁺	−2·44	0·878	−2·44	0·921	−2·48

REFERENCES

Barton, D. E. (1953). *Skand. AktuarTidskr.* **36**, 24.

Barton, D. E. & David, F. N. (1956*a*). *J. R. Statist. Soc.* B (in the Press).

Barton, D. E. & David, F. N. (1956*b*). *Mathematika*, **2**, 150.

David, F. N. & Kendall, M. G. (1955). *Biometrika*, **42**, 223.

Greenwood, M. (1946). *J. R. Statist. Soc.* **109**, 85.

Hall, P. (1927). *Biometrika*, **19**, 240.

Hotelling, H. & Pabst, M. R. (1936). *Ann. Math. Statist.* **7**, 29.

Irwin, J. O. (1927). *Biometrika*, **19**, 225.

Moran, P. A. P. (1947). *Suppl. J. R. Statist. Soc.* **9**, 92.

Olds, C. G. (1938). *Ann. Math. Statist.* **9**, 133.

Pearson, E. S. & Hartley, H. O. (1954). *Biometrika Tables for Statisticians*, **1** Cambridge University Press.

PAIRED COMPARISON DESIGNS FOR TESTING CONCORDANCE BETWEEN JUDGES†

By R. C. BOSE

University of North Carolina and Division of Research Technique, London School of Economics and Political Science

1. Summary and introduction

In a recent paper, 'Further contributions to the theory of paired comparisons', M. G. Kendall (1955) considers paired comparison designs, in which each pair of judges have certain comparisons in common. Such designs should prove useful for testing concordance between judges. He notes that designs of an optimum kind which balance by numbers of comparisons, objects compared, numbers of observers on given comparisons and so forth are rather rare. It is the object of this paper to obtain some paired comparison designs which have a high degree of symmetry. These designs have been defined in §2, and certain inequalities between the parameters are obtained in §3. In §§4 and 5, two special classes of these designs have been investigated, and explicit designs for small values of n (the number of objects to be compared) have been given in Tables 1 and 2. The method of analysis would, to a certain extent, depend on what use the experimenter wants to make of the design. This question will be considered in a future communication.

My thanks are due to Professor Kendall for suggesting the problem, and for helpful discussion during the preparation of the paper.

2. Definition of linked paired comparison designs

Suppose it is required to compare n objects, by employing v judges. Each judge compares r pairs of objects ($r > 1$), and in respect of each pair expresses his opinion whether he prefers one or the other object of the pair. In certain circumstances it may be desirable to allow the judge to express no preference with respect to either of the objects forming the pair. In this case we say that the preference is equally shared by the two objects. We shall assume that the pairs compared by any judge are all different. To ensure symmetry between objects and judges, we require that:

(a) Among the r pairs, compared by each judge, each object appears equally often, say α times.

(b) Each pair is compared by k judges, $k > 1$.

(c) Given any two judges there are exactly λ pairs which are compared by both judges.

Designs satisfying these conditions may be called *linked paired comparison designs.*

Clearly
$$r = \tfrac{1}{2}n\alpha. \tag{2·1}$$

The number of possible pairs is
$$b = \tfrac{1}{2}n(n-1). \tag{2·2}$$

There is a certain correspondence between linked paired comparison designs and balanced incomplete block designs. Each judge may be considered to correspond to a treatment, and

† This research was jointly supported by the United States Air Force through the Office of Scientific Research of the Air Research and Development Command, and the Division of Research Techniques, London School of Economics and Political Science.

each pair to a block. If a pair is compared by a judge, then the block corresponding to the pair may be considered to contain the treatment corresponding to the judge. Hence if a linked paired comparison design of the type considered exists there must exist a corresponding balanced incomplete block design with v treatments and $b = \frac{1}{2}n(n-1)$ blocks, such that each block contains k treatments, each treatment occurs in r blocks, and two given treatments occur together in λ blocks. It follows at once that

$$bk = vr, \quad \lambda(v-1) = r(k-1). \tag{2·3}$$

Using (2·1) and (2·2), the first of the relations (2·3) can be written as

$$v\alpha = k(n-1). \tag{2·4}$$

Also Fisher's well-known inequality (Fisher, 1940; Bose, 1949) gives

$$b \geqslant v \quad \text{or} \quad r \geqslant k. \tag{2·5}$$

Therefore

$$n\alpha \geqslant 2k. \tag{2·6}$$

It should be remembered that the existence of a balanced incomplete block design with parameters v, $b = \frac{1}{2}n(n-1)$, r, k, λ does not ensure the existence of a corresponding linked paired comparison design due to the additional restriction (a). Clearly

$$r \geqslant \lambda. \tag{2·7}$$

The case $r = \lambda$ is trivial, since in this case all the r pairs compared by any judge must also be compared by every other judge. This means that each judge compares the same pairs. Condition (b) then shows that there must be exactly k judges, and each judge compares every pair so that $r = \frac{1}{2}n(n-1)$.

3. Some inequalities

From (2·1) and (2·4)

$$v = \frac{k}{\alpha}\left(\frac{2r}{\alpha} - 1\right).$$

Hence for (2·3)

$$\lambda = \frac{\alpha^2 r(k-1)}{k(2r-\alpha)-\alpha^2} = \frac{\alpha^2}{2} + \frac{\alpha^2(k\alpha+\alpha^2-2r)}{2(2kr-k\alpha-\alpha^2)}. \tag{3·1}$$

This shows that for a given k, λ decreases as r increases. Now $r \geqslant \lambda$, and when $r = \lambda$, we get from (3·1),

$$r = \lambda = \tfrac{1}{2}\alpha(\alpha+1).$$

This proves that

$$r \geqslant \tfrac{1}{2}\alpha(\alpha+1), \quad \lambda \leqslant \tfrac{1}{2}\alpha(\alpha+1), \tag{3·2}$$

where the equality holds only in the trivial case $r = \lambda$.

Since λ must be a positive integer, (3·2) shows that $\alpha = 1$ must imply $r = \lambda = \alpha = 1$. Each judge therefore compares only one pair. Hence the case $\alpha = 1$ is impossible except in the trivial case when there are only two objects and each judge compares them.

Again we can write (3·1) as

$$\lambda = \frac{(k-1)\alpha^2}{2k} + \frac{\alpha^3(k-1)(k+\alpha)}{2k(2kr-k\alpha-\alpha^2)}. \tag{3·3}$$

It follows from (3·2) that

$$2kr - k\alpha - \alpha^2 \geqslant (k-1)\alpha^2.$$

Hence the second term in (3·3) is positive. We thus have the inequality

$$\lambda > \tfrac{1}{2}(k-1)\alpha^2/k. \tag{3·4}$$

Since $k \geqslant 2$, combining (3·4), (3·2) and (3·1), we have

$$\tfrac{1}{4}\alpha^2 < \lambda \leqslant \tfrac{1}{2}\alpha(\alpha+1), \tag{3·5}$$

$$\frac{d\lambda}{dk} = \frac{\alpha^2 r(2r-\alpha-\alpha^2)}{(2kr-k\alpha-\alpha^2)^2}. \tag{3·6}$$

It follows from (3·2) that $$d\lambda/dk \geqslant 0. \tag{3·7}$$

Hence λ is a monotonically increasing function of k, for a given r.

4. Linked paired comparison designs with $\alpha = 2$

The inequality (3·5) shows that when $\alpha = 2$ (neglecting the trivial case $r = \lambda$), we must have $\lambda = 2$. Using (2·1), (2·2), (2·3) and (2·4), we see that all the parameters of the design can be expressed in terms of n, the number of objects to be compared. In fact we have

$$v = \tfrac{1}{2}(n-1)(n-2), \quad b = \tfrac{1}{2}n(n-1), \quad r = n, \quad k = n-2, \quad \lambda = 2, \quad \alpha = 2. \tag{4·1}$$

The existence of (4·1) implies the existence of the balanced incomplete design with parameters

$$v = \tfrac{1}{2}(n-1)(n-2), \quad b = \tfrac{1}{2}n(n-1), \quad r = n, \quad k = n-2, \quad \lambda = 2. \tag{4·2}$$

The balanced incomplete block design (4·2) is known to exist for the values $n = 4$, 5, 6 and 9 (Fisher & Yates, 1938; Bose, 1939), and can in fact be derived by first writing down a solution of the symmetrical balanced incomplete block design,

$$v = b = \tfrac{1}{2}n(n-1)+1, \quad r = k = n, \quad \lambda = 2, \tag{4·3}$$

and then deleting one block and all the treatments in this block. This results in two treatments being deleted from each of the other $\frac{1}{2}n(n-1)$ blocks, since any block of (4·3) has exactly $\lambda = 2$ treatments common with every other block. Conversely, it is known that the existence of (4·2) implies the existence of (4·3) (Connor, 1952; Connor & Hall, 1954). Hence the non-existence of (4·3) for any value of n implies the non-existence of (4·2) for the same values of n. Certain sufficient conditions for the non-existence of (4·3) are known (Shrikhande, 1950; Chowla & Ryser, 1950). In particular, the cases $n = 7$, 8 and 10 are impossible. The design (4·2) is called the derived of (4·3).

We shall now show that for any solution of the symmetrical balanced incomplete block design (4·3), we can derive a corresponding solution of the linked paired comparison design (4·1). The process of derivation will first be illustrated by considering the special case $n = 6$.

It is known (Carmichael, 1937), that a solution for (4·3) in the special case $n = 6$ can be obtained by writing down sixteen treatments in the cells of a 4×4 square and then taking for blocks the six treatments which occur in the same row or the same column as a given treatment (but not including in the block this treatment itself). We shall take the sixteen treatments to be 0, 1, 2, 3, 4, 5, 6 and $a, b, c, d, e, f, g, h, k$ and arrange them as shown below:

0	1	2	3
4	a	b	c
5	d	e	f
6	g	h	k

(4·4)

Then the 16 blocks of (4·3) for the case $n = 6$, are given by

1,	2,	3,	4,	5,	6
1,	2,	0,	c,	f,	k
1,	3,	0,	b,	e,	h
1,	4,	b,	c,	d,	g
1,	5,	a,	g,	e,	f
1,	6,	a,	d,	h,	k
2,	3,	0,	a,	d,	g
2,	4,	a,	c,	e,	h
2,	5,	b,	h,	d,	f
2,	6,	e,	b,	g,	k
3,	4,	a,	b,	f,	k
3,	5,	c,	k,	d,	e
3,	6,	c,	f,	g,	h
4,	5,	0,	g,	h,	k
4,	6,	0,	d,	e,	f
5,	6,	0,	a,	b,	c

(4·5)

A solution of the derived design (4·2) for the case $n = 6$ is now obtainable from (4·5) by omitting the first block in (4·5) and the treatments 1, 2, 3, 4, 5, 6 from the remaining blocks. Thus the derived design is the part included within the lines in (4·5). We can make a (1, 1) correspondence between the blocks of the derived design and the unordered pairs (i, j), $i, j = 1, 2, 3, 4, 5, 6$; $i \neq j$, since each of these blocks has been obtained by deleting just one such pair, from the corresponding block of (4·3). Thus the block $0, c, f, k$ corresponds to the pair (1, 2) and so on. Now let us identify the ten treatments $0, a, b, c, d, e, f, g, h$ and k of the derived with 10 judges, and assign to each judge the pairs corresponding to the blocks in which the treatment corresponding to the judge occurs. We thus get the linked paired comparison design no. (3) of Table 1. Since each of the treatments $0, a, b, c, d, e, f, g, h, k$ occurs in exactly $r = 6$ blocks of the derived, each judge is assigned exactly six pairs, and since any pair of treatments occurs in exactly $\lambda = 2$ blocks, any two judges have just two pairs in common. Again, since each pair of the derived has exactly $k = 4$ treatments, each pair is compared by just 4 judges. Finally, let i be one of the six deleted treatments 1, 2, 3, 4, 5, 6 and let x stand for one of the ten treatments of the derived. Then i and x occur together in only two blocks of the symmetrical design, for example, 1 and a occur together only in the blocks $(1, 5, a, g, e, f)$ and $(1, 6, a, d, h, k)$. Hence the object i occurs twice among the pairs compared by the judge x. This shows that each of the six objects 1, 2, 3, 4, 5, 6 occurs twice among the pairs compared by a judge.

The same process can be used for obtaining a solution of (4·1) for any value of n, for which a solution of the symmetrical balanced incomplete block design (4·3) is known. We first rename the treatments in such a way that the first block contains the treatments $1, 2, \ldots, n$. Then each of the remaining $\frac{1}{2}n(n-1)$ blocks contains just one of the unordered pairs (i,j), $i,j = 1,2,\ldots,n$, $i \neq j$. The remaining $\frac{1}{2}[(n-1)(n-2)]$ treatments may then be identified with judges. If x be one of these treatments, then the judge x is assigned pairs (i,j) corresponding to those blocks in which x occurs.

Cyclic solutions to the symmetrical balanced incomplete block designs (4·3) are known for the cases $n = 4$, 5 and 9. All blocks that can be obtained by developing a suitable initial block mod v. These initial blocks are given blow

n	Initial block	
4	(0, 3, 5, 6)	mod 7
5	(1, 4, 5, 9, 3)	mod 11
9	(1, 16, 34, 26, 9, 33, 10, 12, 7)	mod 37

The corresponding linked paired comparison designs are the designs nos. (1), (2) and (4) of Table 1.

5. Some other types of linked paired designs

Let the number of objects n be even, say $n = 2t$. Then we can divide the $t(2t-1)$ pairs into $2t-1$ sets of t pairs each, such that each object occurs exactly once among the pairs of a set. For example, if $n = 8$, we can take the objects to be 0, 1, 2, 3, 4, 5, 6 and ∞. Then the seven sets are

Sets	Pairs	
I	(1, 6), (2, 5), (3, 4), (0, ∞)	
II	(2, 0), (3, 6), (4, 5), (1, ∞)	
III	(3, 1), (4, 0), (5, 6), (2, ∞)	
IV	(4, 2), (5, 1), (6, 0), (3, ∞)	(5·1)
V	(5, 3), (6, 2), (0, 1), (4, ∞)	
VI	(6, 4), (0, 3), (1, 2), (5, ∞)	
VII	(0, 5), (1, 4), (2, 3), (6, ∞)	

In the general case, the $2t-1$ sets can be obtained by developing mod $(2t-1)$, the initial set

$$(1, 2t-2), \quad (2, 2t-3), \ldots, (t-1, t), \quad (0, \infty) \tag{5·2}$$

the object ∞ remaining unchanged.

Let us now take a balanced incomplete block design with v^* treatments, $b^* = 2t-1$ blocks, r^* replications, block size k^* and in which every pair of treatments occurs together in the same block λ^* times, and make each block correspond to one set and each treatment correspond to one judge. We can then obtain a linked paired comparison design by assigning to each judge the sets of pairs corresponding to all blocks in which the treatment corresponding to the judge occurs. We obtain in this way a linked paired comparison design with parameters

$$n = 2t, \quad v = v^*, \quad b = t(2t-1), \quad r = tr^*, \quad k = k^*, \quad \lambda = t\lambda^*, \quad \alpha = r^*. \tag{5·3}$$

For example, in the case $n = 8$, we may start with the balanced incomplete block design with parameters $v^* = 7$, $b^* = 7$, $r^* = 3$, $k^* = 3$, $\lambda^* = 1$.

If the treatments are taken as a, b, c, d, e, f, g, then the blocks are

$$\begin{array}{lll} a, & b, & d \\ b, & c, & e \\ c, & d, & f \\ d, & e, & g \\ e, & f, & a \\ f, & g, & b \\ g, & a, & c \end{array} \qquad (5{\cdot}4)$$

If we make them correspond to the sets I–VII given by (5·1), then we get the following linked paired comparison design:

Judge	Sets of pairs
a	I, V, VII
b	II, VI, I
c	III, VII, II
d	IV, I, III
e	V, II, IV
f	VI, III, V
g	VII, IV, VI

(5·5)

Table 1

No.	Parameters	Design	
		Judge	Pairs assigned to a judge
(1)	$n = 4,\ v = 3$ $b = 6,\ r = 4$ $k = 2,\ \lambda = 2$ $\alpha = 2$	a b c	(1, 4), (1, 3), (2, 4), (2, 3) (1, 3), (2, 4), (1, 2), (3, 4) (1, 4), (1, 2), (2, 3), (3, 4)
(2)	$n = 5,\ v = 6$ $b = 10,\ r = 5$ $k = 3,\ \lambda = 2$ $\alpha = 2$	a b c d e f	(3, 5), (2, 4), (1, 3), (1, 4), (2, 5) (2, 3), (3, 4), (1, 4), (1, 5), (2, 5) (2, 3), (3, 5), (1, 2), (4, 5), (1, 4) (3, 5), (1, 2), (3, 4), (2, 4), (1, 5) (1, 2), (3, 4), (4, 5), (1, 3), (2, 5) (2, 3), (4, 5), (2, 4), (1, 3), (1, 5)
(3)	$n = 6,\ v = 10$ $b = 15,\ r = 6$ $k = 4,\ \lambda = 2$ $\alpha = 2$	0 a b c d e f g h k	(1, 2), (1, 3), (2, 3), (4, 5), (4, 6), (5, 6) (2, 3), (2, 4), (3, 4), (1, 5), (1, 6), (5, 6) (1, 3), (1, 4), (3, 4), (2, 5), (2, 6), (5, 6) (1, 4), (2, 4), (1, 2), (3, 5), (3, 6), (5, 6) (1, 4), (1, 6), (4, 6), (2, 3), (2, 5), (3, 5) (1, 3), (1, 5), (3, 5), (2, 4), (2, 6), (4, 6) (1, 2), (1, 5), (2, 5), (3, 4), (3, 6), (4, 6) (1, 4), (1, 5), (4, 5), (2, 3), (2, 6), (3, 6) (1, 3), (1, 6), (3, 6), (2, 4), (2, 5), (4, 5) (1, 2), (1, 6), (2, 6), (3, 4), (3, 5), (4, 5)

Table 1 (*cont.*)

Linked paired designs with $\alpha = 2$			
No.	Parameters	Design	
		Judge	Pairs
(4)	$n = 9,\ v = 28$ $b = 36,\ r = 9$ $k = 7,\ \lambda = 2$ $\alpha = 2$	a	(5, 6), (6, 7), (7, 8), (8, 9), (1, 9), (1, 2), (2, 3), (3, 4), (4, 5)
		b	(3, 7), (6, 8), (1, 8), (5, 7), (4, 5), (2, 9), (1, 4), (2, 3), (6, 9)
		c	(5, 8), (3, 6), (2, 9), (4, 7), (1, 7), (2, 6), (4, 5), (1, 9), (3, 8)
		d	(7, 8), (3, 4), (2, 6), (2, 8), (6, 9), (3, 5), (1, 7), (1, 4), (5, 9)
		e	(1, 2), (1, 6), (3, 5), (4, 8), (3, 8), (2, 7), (6, 9), (4, 5), (7, 9)
		f	(1, 8), (2, 3), (2, 7), (4, 6), (5, 9), (4, 9), (3, 8), (1, 7), (5, 6)
		g	(2, 6), (1, 4), (8, 9), (2, 4), (5, 6), (1, 5), (7, 9), (3, 8), (3, 7)
		h	(4, 9), (6, 9), (4, 7), (1, 3), (5, 8), (2, 8), (3, 7), (5, 6), (1, 2)
		i	(1, 5), (5, 9), (4, 8), (3, 6), (1, 2), (4, 6), (7, 8), (3, 7), (2, 9)
		j	(5, 7), (7, 9), (4, 6), (3, 4), (1, 8), (3, 9), (1, 2), (5, 8), (2, 6)
		k	(4, 7), (5, 6), (3, 9), (1, 6), (2, 9), (2, 4), (1, 8), (7, 8), (3, 5)
		l	(4, 8), (3, 7), (2, 5), (1, 9), (3, 5), (6, 7), (2, 6), (1, 8), (4, 9)
		m	(4, 6), (5, 8), (6, 7), (1, 4), (2, 7), (1, 3), (3, 5), (2, 9), (8, 9)
		n	(3, 9), (7, 8), (1, 3), (4, 5), (4, 9), (6, 8), (2, 7), (2, 6), (1, 5)
		o	(2, 4), (1, 2), (6, 8), (1, 7), (8, 9), (3, 6), (4, 9), (3, 5), (5, 7)
		p	(2, 5), (1, 8), (3, 6), (6, 9), (1, 5), (3, 4), (8, 9), (2, 7), (4, 7)
		q	(6, 7), (2, 9), (3, 4), (3, 8), (5, 7), (1, 6), (1, 5), (4, 9), (2, 8)
		r	(1, 3), (2, 6), (1, 6), (5, 9), (4, 7), (2, 3), (5, 7), (8, 9), (4, 8)
		s	(6, 8), (3, 5), (2, 3), (7, 9), (2, 8), (1, 9), (4, 7), (1, 5), (4, 6)
		t	(3, 6), (2, 7), (1, 9), (5, 6), (4, 8), (1, 4), (2, 8), (5, 7), (3, 9)
		u	(1, 6), (8, 9), (4, 5), (3, 7), (3, 9), (1, 7), (4, 6), (2, 8), (2, 5)
		v	(2, 3), (1, 5), (1, 7), (5, 8), (2, 4), (6, 9), (3, 9), (4, 8), (6, 7)
		w	(1, 9), (5, 7), (6, 9), (7, 8), (2, 5), (3, 8), (2, 4), (4, 6), (1, 3)
		x	(1, 4), (4, 7), (3, 8), (1, 2), (6, 7), (5, 9), (2, 5), (3, 9), (6, 8)
		y	(4, 5), (2, 8), (5, 9), (1, 8), (1, 3), (7, 9), (6, 7), (2, 4), (3, 6)
		z	(1, 7), (4, 8), (7, 9), (2, 9), (6, 8), (5, 6), (1, 3), (2, 5), (3, 4)
		α	(5, 9), (2, 4), (3, 7), (2, 7), (1, 6), (5, 8), (3, 4), (6, 8), (1, 9)
		β	(7, 9), (2, 5), (5, 8), (4, 9), (2, 3), (7, 8), (1, 6), (3, 6), (1, 4)

The parameters of the design are

$$n = 8, \quad v = 7, \quad b = 28, \quad r = 12, \quad k = 3, \quad \lambda = 4, \quad \alpha = 3.$$

Each judge compares 12 of the 28 possible pairs. Other designs obtained in this way are the designs nos. (1), (2) and (4) of Table 2. It should be noticed that the design (1) of Table 2 is the same as the design (1) of Table 1, obtained in a different manner.

Again let the number of objects be odd, say $n = 2t+1$. Then we can divide the $t(2t+1)$ pairs into t sets of $(2t+1)$ pairs each, such that each object occurs exactly twice among the pairs of a set. For example, if $n = 7$, we can take the objects to be 0, 1, 2, 3, 4, 5, 6. Then the three sets are:

Sets	Pairs	
I	(0, 1), (1, 2), (2, 3), (3, 4), (4, 5), (5, 6), (6, 0)	
II	(0, 2), (1, 3), (2, 4), (3, 5), (4, 6), (5, 0), (6, 1)	(5·6)
III	(0, 3), (1, 4), (2, 5), (3, 6), (4, 0), (5, 1), (6, 2)	

In the general case we can take the objects to be $0, 1, \ldots, 2t$. Then the ith set consists of all pairs for which the difference mod $(2t+1)$ between the objects constituting the pair is i.

Let us now take a balanced incomplete block design with v' treatments, $b' = t$ blocks, r' replications, block size k' and in which every pair of treatments occurs together in the same block λ' times, and make each block correspond to one set and each treatment to a judge. We then get a linked paired comparison design by assigning to each judge the sets of pairs corresponding to all blocks, in which the treatment corresponding to the judge occurs. We obtain in this way a linked paired design with parameters

$$n = 2t+1, \quad v = v', \quad b = t(2t+1), \quad r = (2t+1)r', \quad k = k', \quad \lambda = (2t+1)\lambda', \quad \alpha = 2r'. \quad (5{\cdot}7)$$

For example, in the case $n = 7$, we get the linked paired comparison design no. (5) of Table 2. It should be noted that the sets obtained in this case are the same as the tours round the preference polygon considered by Kendall (1955), and lead to the designs already considered by him.

Table 2

No.	Parameters	Sets of pairs	Design	
			Judge	Sets of pairs assigned to a judge
(1)	$n=4,\ v=3$ $b=6,\ r=4$ $k=2,\ \lambda=2$ $\alpha=2$	I (1, 2), (0, ∞) II (2, 0), (1, ∞) III (0, 1), (2, ∞)	*a* *b* *c*	II, III III, I I, II
(2)	$n=6,\ v=5$ $b=15,\ r=12$ $k=4,\ \lambda=9$ $\alpha=4$	I (1, 4), (2, 3), (0, ∞) II (2, 0), (3, 4), (1, ∞) III (3, 1), (4, 0), (2, ∞) IV (4, 2), (0, 1), (3, ∞) V (0, 3), (1, 2), (4, ∞)	*a* *b* *c* *d* *e*	II, III, IV, V I, III, IV, V I, II, IV, V I, II, III, V I, II, III, IV
(3)	$n=8,\ v=7$ $b=28,\ r=12$ $k=3,\ \lambda=4$ $\alpha=3$	The sets (5·1)	The design (5·5)	
(4)	$n=8,\ v=7$ $b=28,\ r=16$ $k=4,\ \lambda=8$ $\alpha=4$	The sets (5·1)	*a* *b* *c* *d* *e* *f* *g*	II, V, VI, VII IV, VI, VII, I V, VII, I, II VI, I, II, III VII, II, III, IV I, III, IV, V II, IV, V, VI
(5)	$n=7,\ v=3$ $b=21,\ r=14$ $k=2,\ \lambda=7$ $\alpha=4$	The sets (5·6)	*a* *b* *c*	II, III III, I I, II

REFERENCES

Bose, R. C. (1939). On the construction of balanced incomplete block designs. *Ann. Eugen., Lond.*, **9**, 353–99.

Bose, R. C. (1949). A note on Fisher's inequality for balanced incomplete block designs. *Ann. Math. Statist.* **20**, 619–20.

Carmichael, R. D. (1937). *Introduction to the Theory of Groups of Finite Order*. Boston and London: Ginn and Co.

Chowla, S. & Ryser, H. T. (1950). Combinatorial problems. *Canad. J. Math.* **2**, 93–9.

Connor, W. S. (1952). On the structure of balanced incomplete block designs. *Ann. Math. Statist.* **23**, 57–71.

Connor, W. S. & Hall, M. (1954). An embedding theorem for balanced incomplete block designs. *Canad. J. Math.* **6**, 35–41.

Fisher, R. A. (1940). An examination of the different possible solutions of a problem in incomplete blocks. *Ann. Eugen., Lond.*, **10**, 52–75.

Fisher, R. A. & Yates, F. (1938). *Statistical Tables for Biological, Agricultural and Medical Research*. London and Edinburgh: Oliver and Boyd.

Kendall, M. G. (1955). Further contributions to the theory of paired comparisons. *Biometrics*, **11**, 43–62.

Shrikhande, S. S. (1950). The impossibility of certain symmetrical balanced incomplete block designs. *Ann. Math. Statist.* **21**, 106–11.

ON THE DISTRIBUTION OF THE LARGEST OR THE SMALLEST ROOT OF A MATRIX IN MULTIVARIATE ANALYSIS*

By K. C. S. PILLAI†

University of North Carolina and University of Travancore

1. Introduction

This paper deals with the distribution problem of the extreme characteristic roots of a matrix in multivariate analysis (Roy, 1939; Hsu, 1939; Fisher, 1939). Roy (1945, 1953, 1954) has discussed in detail the usefulness of the extreme characteristic roots in tests of hypotheses and confidence interval estimation in multivariate situations.

For obtaining tests of three different types of hypotheses in multivariate analysis, namely, (i) that of equality of the dispersion matrices of two p-variate normal populations, (ii) that of equality of the p-dimensional mean vectors for l p-variate normal populations and (iii) that of independence between a p-set and a q-set ($p \leqslant q$) of variates in a $(p+q)$-variate normal population, in each case we arrive at the characteristic roots of a matrix based on sample observations. In each case, if the hypothesis to be tested is true, the non-zero roots $(0 < \theta_1 \leqslant \theta_2 \leqslant \ldots \leqslant \theta_s < 1;\ s \leqslant p)$ have the same joint distribution, the form of which was given by Roy (1939), Hsu (1939) and Fisher (1939). The distribution can be written in the form

$$p(\theta_1, \ldots, \theta_s) = C(s, m, n) \prod_{i=1}^{s} \theta_i^m (1-\theta_i)^n \prod_{i>j} (\theta_i - \theta_j) \quad (0 < \theta_1 \leqslant \ldots \leqslant \theta_s < 1), \tag{1}$$

where

$$C(s, m, n) = \pi^{\frac{1}{2}s} \prod_{i=1}^{s} \Gamma\left(\frac{2m+2n+s+i+2}{2}\right) \Big/ \prod_{i=1}^{s} \left\{\Gamma\left(\frac{2m+i+1}{2}\right) \Gamma\left(\frac{2n+i+1}{2}\right) \Gamma(\tfrac{1}{2}i)\right\}, \tag{2}$$

and where m and n have to be interpreted differently for the different situations (Pillai, 1954, 1955).

The problem of obtaining the cumulative distribution function (c.d.f.) of the largest root has been already investigated by Roy (1945), who gave explicit expressions for the c.d.f. for numbers of roots 2, 3 and 4. Nanda (1948) also gave such expressions for $s = 2$, 3, 4 and 5. The present author (1954) has extended these results for the c.d.f. for values of s up to 8. The expressions involved in these results are functions of incomplete beta functions and are useful for extensive tabulation of the probability integral. However, if one considers only the problem of obtaining the upper percentage points (5 % or less), the approximations suggested in the following section will be useful.

2. Approximate formulae for obtaining upper percentage points (5 % or less) of the largest root

Let us first consider the case of two roots. The c.d.f. of the largest root in this case involves only two incomplete beta functions and can be written in an explicit form (Nanda, 1948; Pillai, 1954) given by

$$\Pr(\theta_2 \leqslant x) = \frac{C(2, m, n)}{(m+n+2)} \left\{2 \int_0^x \theta_2^{2m+1} (1-\theta_2)^{2n+1} d\theta_2 - x^{m+1}(1-x)^{n+1} \int_0^x \theta_2^m (1-\theta_2)^n d\theta_2\right\}. \tag{3}$$

* Part of a doctoral thesis, submitted to the Department of Statistics, University of North Carolina, Chapel Hill; work done under the sponsorship of the Ford Foundation.

† Now with United Nations, New York.

For integral values of m, integration by parts of the two integrals in (3) will reduce the probability integral in (3) to the form

$$\Pr(\theta_2 \leqslant x) = \frac{C(2,m,n)}{(m+n+2)}\Big\{\sum_{i=0}^{m}\frac{(m)_i\,\Gamma(n+1)\,x^{2m-i+1}(1-x)^{2n+i+2}}{\Gamma(n+i+2)} - 2\sum_{i=0}^{2m+1}\frac{(2m+1)_i\,\Gamma(2n+2)\,x^{2m-i+1}(1-x)^{2n+i+2}}{\Gamma(2n+i+3)} + \frac{2\Gamma(2m+2)\,\Gamma(2n+2)}{\Gamma(2m+2n+4)} - \frac{\Gamma(m+1)\,\Gamma(n+1)}{\Gamma(m+n+2)}x^{m+1}(1-x)^{n+1}\Big\}, \quad (4)$$

where $(t)_r = t(t-1)\ldots(t-r+1)$.

In (4) if we neglect terms of order higher than $x^{m+1}, (1-x)^{n+1}$, we get

$$\Pr(\theta_2 \leqslant x) = \frac{C(2,m,n)}{(m+n+2)}\Big\{\frac{2\Gamma(2m+2)\,\Gamma(2n+2)}{\Gamma(2m+2n+4)} - \frac{\Gamma(m+1)\,\Gamma(n+1)}{\Gamma(m+n+2)}x^{m+1}(1-x)^{n+1}\Big\}. \quad (5)$$

For small values of m, (5) will give the percentage points for the upper 5 % level or less. Using the fact that $\Pr(\theta_2 \leqslant 1) = 1$, we obtain (5) in the simpler form

$$1 - \frac{\Gamma(m+1)\,\Gamma(n+1)\,\Gamma(2m+2n+4)}{2\Gamma(m+n+2)\,\Gamma(2m+2)\,\Gamma(2n+2)}x^{m+1}(1-x)^{n+1}. \quad (6)$$

If the probability in (6) has to be 0·95, then

$$\frac{\Gamma(m+1)\,\Gamma(n+1)\,\Gamma(2m+2n+4)}{2\Gamma(m+n+2)\,\Gamma(2m+2)\,\Gamma(2n+2)}x^{m+1}(1-x)^{n+1} = 0{\cdot}05. \quad (7)$$

The expression in (7) is very simple for computation, especially when the values of m are small. The error involved in using this approximation has been computed and difference between the exact and the approximate probabilities has been found to occur in the fifth decimal place, and, on rounding off, there could be at the most a difference of 1 in the fourth place.

Proceeding on the same lines as above, i.e. as for the approximation in (6), we get the following approximations for $s = 3$, 4 and 5:

$$\Pr(\theta_3 \leqslant x) = \frac{\Gamma(m+n+3)}{\Gamma(m+2)\,\Gamma(n+2)}\Big[\Big\{\int_0^x \theta_3^m(1-\theta_3)^n\,d\theta_3\Big\}(m+1)(2m+3) - \Big\{\int_0^x \theta_3^{m+1}(1-\theta_3)^n\,d\theta_3\Big\}(m+1)(2m+2n+5) - x^{m+2}(1-x)^{n+1}(2m+2n+5)\Big], \quad (8)$$

$$\Pr(\theta_4 \leqslant x) = 1 - \frac{\Gamma(2m+2n+8)\,\Gamma(m+2)\,\Gamma(n+2)}{2\Gamma(m+n+3)\,\Gamma(2m+4)\,\Gamma(2n+4)}x^{m+3}(1-x)^{n+1} + \frac{\Gamma(m+1)\,\Gamma(n+2)\,\Gamma(2m+2n+8)}{4\Gamma(2m+2)\,\Gamma(2n+4)\,\Gamma(m+n+4)}x^{m+2}(1-x)^{n+1} - \frac{(m+2)\,\Gamma(m+1)\,\Gamma(n+2)\,\Gamma(2m+2n+7)}{4\Gamma(2m+2)\,\Gamma(2n+4)\,\Gamma(m+n+4)}x^{m+1}(1-x)^{n+1}, \quad (9)$$

$$\Pr(\theta_5 \leqslant x) = \frac{\Gamma(m+n+4)}{3\Gamma(m+1)\,\Gamma(n+3)}\Big\{(2m+3)\,(2m+5)\int_0^x \theta_5^m(1-\theta_5)^n\,d\theta_5$$
$$+\frac{(2m+5)\,(2m+2n+7)\,(2m^2+2mn+13m+12)}{(2n+3)}\int_0^x \theta_5^{m+1}(1-\theta_5)^n\,d\theta_5$$
$$-\frac{(2m+3)\,(2m+2n+7)\,(2m+2n+9)\,(m+n+4)}{(2n+3)}\int_0^x \theta_5^{m+2}(1-\theta_5)^n\,d\theta_5$$
$$-\frac{(m+3)\,(2m+5)\,(2m+2n+5)\,(2m+2n+7)}{(2n+3)\,(m+1)}x^{m+2}(1-x)^{n+1}$$
$$+3\frac{(2m+2n+7)\,(2m+2n+9)}{(m+1)}x^{m+3}(1-x)^{n+1}$$
$$-\frac{(2m+2n+7)\,(2m+2n+9)\,(m+n+4)}{(m+1)\,(m+2)}x^{m+4}(1-x)^{n+1}\Big\}. \tag{10}$$

The incomplete beta functions in (8) and (10) can be further integrated by parts, and for small values of m the expressions will become simpler and could be readily used for computation.

3. Percentage points for the largest root

An important problem in multivariate analysis is to test the hypothesis that l p-variate normal populations having the same variance-covariance matrix, from each of which a sample is drawn, have the same mean for each variate. This, in fact, is the second one among the three tests discussed in the introduction of the paper. For this test it turns out that

$$m = \frac{|l-p-1|-1}{2} \quad \text{and} \quad n = \frac{N-p-l-1}{2}. \tag{11}$$

For $p = 2$ we have $m = \frac{1}{2}(l-4)$, and the expressions for the upper percentage points given in § 2 which are good for small values of m become quite useful, because the number of samples, l, cannot be too large. For example, the case of $m = 4$, i.e. $l = 12$, relates to a problem with 12 samples. Using (7) tables of upper 5 and 1 % points have been computed for two roots for values of $m = 0$, 1, 2, 3 and 4 and $n = 5$, 10, 15, 20, 25, 30, 40, 60, 80, 100, 130, 160, 200, 300, 500 and 1000. Significance levels for fractional values of m and intermediate values of n can be obtained by interpolation. The percentage points based on the approximate formula (7) are given in Table 1.

In order to form an idea of the error of approximation, the probability based on the exact formulae is calculated at the percentage points for the parameters $m = 2$ and $n = 10$, 30 and 80 and $m = 4$, $n = 5$ and 100, picked out from Table 1 (based on the approximate formula (7)). The difference between the exact and approximate probabilities is exhibited in Table 2. It is to be noted that within the range of values of the parameters considered, the approximation is quite good.

Nanda (1951) has given significance levels (upper 5 and 1 % for the largest root for $s = 2$ and very small values of m and n ($m = 0(\frac{1}{2})\,2$; $n = \frac{1}{2}(\frac{1}{2})\,10$). His table has been computed by using the *Tables of the Incomplete Beta Function* (1934) for some values of m and n, the other values being obtained by interpolation. The significance levels are given only up to two decimals, and the range of m and n considered is very small. In Table 1, however, the range for n is from 5 to 1000 and for m is 0(1) 4. The significance levels have been given correct to

Table 1. *Percentage points of the largest root for* $s = 2$

(*a*) *Upper* 5 % *points*

n \ m	0	1	2	3	4
5	0·565	0·651	0·706	0·746	0·776
10	0·374	0·455	0·514	0·561	0·598
15	0·278	0·348	0·402	0·445	0·483
20	0·222	0·281	0·329	0·369	0·403
25	0·183	0·236	0·278	0·314	0·346
30	0·157	0·203	0·241	0·274	0·303
40	0·121	0·158	0·190	0·218	0·243
60	0·0836	0·110	0·133	0·154	0·173
80	0·0638	0·0846	0·1027	0·1191	0·1345
100	0·0515	0·0686	0·0835	0·0972	0·1100
130	0·0400	0·0535	0·0652	0·0761	0·0864
160	0·0327	0·0437	0·0535	0·0626	0·0711
200	0·0263	0·0352	0·0432	0·0506	0·0576
300	0·0176	0·0237	0·0291	0·0342	0·0390
500	0·0106	0·0143	0·0176	0·0207	0·0237
1000	0·00535	0·00719	0·00888	0·01045	0·01195

(*b*) *Upper* 1 % *points*

n \ m	0	1	2	3	4
5	0·675	0·745	0·787	0·817	0·839
10	0·470	0·544	0·597	0·638	0·670
15	0·357	0·425	0·476	0·517	0·551
20	0·288	0·347	0·394	0·433	0·467
25	0·240	0·293	0·336	0·372	0·403
30	0·207	0·254	0·293	0·326	0·355
40	0·161	0·200	0·232	0·261	0·286
60	0·1114	0·140	0·165	0·186	0·206
80	0·0852	0·1080	0·1273	0·1448	0·1609
100	0·0692	0·0878	0·1038	0·1184	0·1319
130	0·0539	0·0685	0·0813	0·0930	0·1039
160	0·0441	0·0562	0·0668	0·0765	0·0857
200	0·0355	0·0453	0·0540	0·0619	0·0694
300	0·0239	0·0305	0·0365	0·0419	0·0471
500	0·0144	0·0185	0·0221	0·0255	0·0287
1000	0·00725	0·00930	0·01114	0·01285	0·01448

three significant figures, in order to ensure sufficient accuracy for interpolation. For testing the equality of variate means from different populations, Table 1 should be enough up to $m = 4$, but if anything beyond this is needed one can obtain the significance level easily from the expression in (6) for small values of m.

Table 2. *The error of approximation at the* 5 % *level* (*i.e. difference in value between* (3) *and* (6) *for* $s = 2$ *at the* 5 % *significance level in Table* 1)

n	$m = 2$ difference (approximate − exact)	n	$m = 4$ difference (approximate − exact)
10	0·00002693	5	0·00002574
30	0·00002697	100	0·00003477
80	0·00002757	—	—

4. The distributions of the largest and smallest roots for $s = 2$ as hypergeometric series

The joint-probability density function of θ_1 and θ_2 is given by

$$p(\theta_1, \theta_2) = C(2, m, n)\,(\theta_1\theta_2)^m\,(1-\theta_1)^n\,(1-\theta_2)^n\,(\theta_2-\theta_1) \quad (\theta \leqslant \theta_1 \leqslant \theta_2 \leqslant 1), \tag{12}$$

where

$$C(2, m, n) = \Gamma(2m + 2n + 5)/\{2^2\Gamma(2m+2)\,\Gamma(2n+2)\}.$$

Now consider the transformation

$$\lambda = \theta_1/\theta_2, \quad \theta_2 = \theta_2. \tag{13}$$

Then
$$p(\lambda, \theta_2) = C(2, m, n)\,\theta_2^{2m+2}(1-\theta_2)^n\,(1-\lambda\theta_2)^n\,\lambda^m(1-\lambda), \tag{14}$$

where the range of variation of λ as well as θ_2 is 0–1. Now integrating out λ from $p(\lambda, \theta_2)$,

$$p(\theta_2) = C(2, m, n)\,\theta_2^{2m+2}(1-\theta_2)^n \int_0^1 (1-\lambda\theta_2)^n\,\lambda^m(1-\lambda)\,d\lambda, \tag{15}$$

or
$$p(\theta_2) = \frac{C(2, m, n)}{(m+1)\,(m+2)}\,\theta_2^{2m+2}(1-\theta_2)^n\,F(m+1, -n, m+3, \theta_2), \tag{16}$$

where

$$F(m+1, -n, m+3, \theta_2) = 1 - \frac{n(m+1)}{m+3}\theta_2 + \frac{n(n-1)\,(m+1)\,(m+2)}{2\,!\,(m+3)\,(m+4)}\theta_2^2 - \ldots. \tag{17}$$

Again starting with (12) and effecting the following transformation

$$(1-\theta_1)\,\mu = 1-\theta_2 \quad \text{and} \quad \theta_1 = \theta_1, \tag{18}$$

we get
$$p(\theta_1) = C(2, m, n)\,\theta_1^m(1-\theta_1)^{2n+2}\int_0^1 \mu^n(1-\mu)\,\{1-\mu(1-\theta_1)\}^m\,d\mu, \tag{19}$$

or
$$p(\theta_1) = \frac{C(2, m, n)}{(n+1)\,(n+2)}\,\theta_1^m(1-\theta_1)^{2n+2}\,F(n+1, -m, n+3, 1-\theta_1). \tag{20}$$

It may be observed that a change from $1-\theta_1$ to θ_2 and m to n in (20) gives (16). Hence

$$\Pr(\theta_2 \leqslant x) = \Pr(\theta_2 \leqslant x;\, m, n) = \Pr(1-\theta_1 \leqslant x;\, n, m) = \Pr(1-x \leqslant \theta_1;\, n, m)$$
$$= 1 - \Pr(\theta_1 < 1-x;\, n, m). \tag{21}$$

Hence lower percentage points for the smallest root can be obtained from the upper percentage points of the largest root when the parameters m and n are interchanged.

The result (21) is true for any value of s as Nanda (1948) has shown, starting from distribution (1). Hence the expressions for the c.d.f. of the largest root given in § 2 can be used to obtain the lower percentage points (5 % or less) of the smallest root by noting the relation (21) and substituting s for 2.

5. SUMMARY

An approximation to the cumulative distribution function of the largest (smallest) root of a matrix in multivariate analysis for obtaining the upper (5 % or less) percentage points for the largest root (lower (5 % or less) percentage points of the smallest) is given for any number of roots up to 5. Percentage points for two roots for small integral values of m have also been presented. The distributions of the largest and smallest roots are also developed as hypergeometric series in the case of $s = 2$.

I wish to acknowledge my indebtedness to Prof. S. N. Roy for his kind advice and criticism in the course of my investigations. I am also indebted to Prof. H. Hotelling for his active interest and valuable suggestions in the preparation of this paper.

REFERENCES

FISHER, R. A. (1939). *Ann. Eugen., Lond.*, **9**, 238.
HSU, P. L. (1939). *Ann. Eugen., Lond.*, **9**, 250.
NANDA, D. N. (1948). *Ann. Math. Statist.* **19**, 340.
NANDA, D. N. (1951). *J. Ind. Soc. Agric. Statist.* **3**, 175.
PEARSON, K. (1934). *Tables of the Incomplete Beta Function*. Cambridge University Press for the Biometrika Trustees.
PILLAI, K. C. S. (1954). *On Some Distribution Problems in Multivariate Analysis*, Mimeo Series no. 88, Institute of Statistics, University of North Carolina.
PILLAI, K. C. S. (1955). *Ann. Math. Statist.* **26**, 117.
ROY, S. N. (1939). *Sankhyā*, **4**, 381.
ROY, S. N. (1945). *Sankhyā*, **7**, 133.
ROY, S. N. (1953). *Ann. Math. Statist.* **24**, 220.
ROY, S. N. (1954). *A Report on Some Aspects of Multivariate Analysis*, Mimeo. Series no. 121, Institute of Statistics, University of North Carolina.

TESTS OF SIGNIFICANCE FOR THE LATENT ROOTS OF COVARIANCE AND CORRELATION MATRICES

By D. N. LAWLEY
Mathematical Institute, Edinburgh

1. Statement of problem

In a recent note Bartlett (1954) has summarized various approximate χ^2 tests (see also other papers referred to therein). Our object in this paper is to extend certain of these, listed as III a, b, c, which may be regarded as tests involving the latent roots of sample covariance and correlation matrices. We shall be interested in those cases where the effects of the k largest latent roots have been removed and where a hypothesis of equality of the remaining roots is made. Such hypotheses are made in principal components and factor analyses.

2. Residual latent roots of a covariance matrix when the true value λ is known

Suppose that $x_1, x_2, \ldots, x_p$ are p variates following a multivariate normal distribution with true covariance matrix $A_0 = [\alpha_{ij}]$. Assuming that the first k latent roots $\lambda_1, \lambda_2, \ldots, \lambda_k$ of A_0, in order of magnitude, are distinct, we make the hypothesis that the residual latent roots $\lambda_{k+1}, \lambda_{k+2}, \ldots, \lambda_p$ are equal to some value λ. We shall consider first the case where λ is known.

Let $A = [a_{ij}]$ be a sample covariance matrix with n degrees of freedom (corresponding as a rule to a sample of size $n+1$). Let l_i $(i = 1, 2, \ldots, p)$ be the latent roots in order of magnitude of A. It is assumed that n is sufficiently large for us to be able to set up an almost certain correspondence between the first k of the l_i and the first k of the λ_i. More precisely, we require that sampling errors, which are of order $1/\sqrt{n}$, should be small compared with the differences between successive quantities $\lambda_1, \lambda_2, \ldots, \lambda_k, \lambda$. The degree of uncertainty in establishing the correspondence between the first k of the l_i and the first k of the λ_i, though difficult to evaluate, appears to decrease very rapidly as n increases and is in general unlikely to affect appreciably the approximations developed in the paper.

To test the above hypothesis we use as an approximate χ^2, with $\frac{1}{2}(p-k)(p-k+1)$ degrees of freedom, the expression

$$-\log_e (l_{k+1} l_{k+2} \ldots l_p / \lambda^{p-k}) + (l_{k+1} + l_{k+2} + \ldots + l_p)/\lambda - (p-k), \tag{1}$$

multiplied by an appropriate factor which we have to determine. (It is easy to verify that for $k = 0$, this is equivalent to Bartlett's test if we put $V_0 = \lambda I$ in his expression of III a.) The multiplying factor will be determined by finding the expectation of expression (1), since this gives a better approximation than the cruder value n.

To simplify the problem we may, without loss of generality, assume that all true correlations are zero, i.e. that A_0 is diagonal. This is justifiable, since any orthogonal transformation of the variates x_i leaves the latent roots of the covariance matrix unaltered. It will be convenient to partition A_0 and A in such a way as to divide the first k rows and columns from the last $(p-k)$ rows and columns. We may then write

$$A_0 = \begin{bmatrix} \Lambda & 0 \\ 0 & \lambda I \end{bmatrix}, \quad A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

where Λ is the diagonal matrix with elements $\lambda_1, \lambda_2, \ldots, \lambda_k$.

We shall use the symbol δ to denote deviations of sample values from expectation so that, for example, $\delta A = A - A_0$, $\delta a_{ij} = a_{ij} - \alpha_{ij}$, $\delta l_i = l_i - \lambda_i$. All such quantities are of order $1/\sqrt{n}$, and in what follows we shall be able to neglect terms of sufficiently high degree in the δ's. Partitioning δA in the same manner as A, we may write

$$A_{11} = \Lambda + \delta A_{11},$$
$$A_{22} = \lambda I + \delta A_{22},$$
$$A_{12} = \delta A_{12}.$$

If we are prepared to neglect the squares of the elements of δA_{12}, the residual latent roots of A could be taken as being approximately equal to the latent roots of A_{22}. This would not, however, give us sufficient accuracy, since we wish to evaluate the expectation of expression (1) as far as terms in $1/n^2$. To accomplish this we define a matrix B having the same latent roots as A and given by

$$B = (I+E)\,A\,(I+E)^{-1},$$

where

$$E = -E' = \begin{bmatrix} E_{11} & E_{12} \\ -E_{12}' & 0 \end{bmatrix},$$

$$E_{12} = (\Lambda - \lambda I)^{-1}\,\delta A_{12},$$

$$E_{11} = \begin{bmatrix} 0 & \dfrac{\delta a_{12}}{\lambda_1 - \lambda_2} & \cdots & \dfrac{\delta a_{1k}}{\lambda_1 - \lambda_k} \\ \dfrac{\delta a_{12}}{\lambda_2 - \lambda_1} & 0 & \cdots & \dfrac{\delta a_{2k}}{\lambda_2 - \lambda_k} \\ \vdots & \vdots & & \vdots \\ \dfrac{\delta a_{1k}}{\lambda_k - \lambda_1} & \dfrac{\delta a_{2k}}{\lambda_k - \lambda_2} & \cdots & 0 \end{bmatrix}.$$

We now write $B = (I+E)\,(A_0 + \delta A)\,(I+E)^{-1}$

$$= A_0 + \{(EA_0 - A_0E + \delta A) + E\delta A\}\,(I - E + E^2 - \ldots),$$

and we note that $EA_0 - A_0E + \delta A = \begin{bmatrix} E_{11}\Lambda - \Lambda E_{11} + \delta A_{11} & 0 \\ 0 & \delta A_{22} \end{bmatrix},$

and that $E_{11}\Lambda - \Lambda E_{11} + \delta A_{11}$ is a diagonal matrix with elements $\delta a_{11}, \delta a_{22}, \ldots, \delta a_{kk}$. Hence, partitioning B in the usual manner, we have

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix},$$

where the elements of B_{12}, B_{21} and the non-diagonal elements of B_{11} are all of order δ^2. From this it follows that with errors of order δ^4 the residual roots $l_{k+1}, l_{k+2}, \ldots, l_p$ are equal to the latent roots of

$$B_{22} = A_{22} - \delta A_{21}(\Lambda - \lambda I)^{-1}\,\delta A_{12} - \delta A_{22}\,\delta A_{21}(\Lambda - \lambda I)^{-2}\,\delta A_{12} + \delta A_{21}(\Lambda - \lambda I)^{-1}\,\delta A_{11}(\Lambda - \lambda I)^{-1}\,\delta A_{12} + O(\delta^4). \quad (2)$$

It also follows that, to the same degree of accuracy, $l_r (r = 1, 2, \ldots, k)$ is equal to the rth diagonal element of B_{11}, which is

$$\lambda_r + \delta a_{rr} + \sum_i{}' \frac{(\delta a_{ri})^2}{\lambda_r - \lambda_i} - \delta a_{rr} \sum_i{}' \left(\frac{\delta a_{ri}}{\lambda_r - \lambda_i}\right)^2 + \sum_i{}' \sum_j{}' \left\{\frac{\delta a_{ri}\,\delta a_{ij}\,\delta a_{jr}}{(\lambda_r - \lambda_i)\,(\lambda_r - \lambda_j)}\right\} + O(\delta^4), \quad (3)$$

where Σ' indicates that the summation is to be over all values from 1 to p except r.

It is easy to see that errors of order δ^4 in the latent roots $l_{k+1}, l_{k+2}, \ldots, l_p$ will give rise to an error only of order δ^5 in expression (1), so that in it we may replace the product of the latent roots by the determinant of B_{22} and the sum by the trace of B_{22}. In the ensuing algebra it will be convenient to put $\lambda = 1$ and to replace λ_r by λ_r/λ at the conclusion. The expression (1) then becomes

$$-\log_e |B_{22}| + \operatorname{tr}(B_{22}) - q,$$

where $q = p - k$.

Now, writing B_{22} as $A_{22} - Z$, where Z is given by (2), we have

$$\begin{aligned} |B_{22}| &= |A_{22} - Z| \\ &= |A_{22}|\,|I - A_{22}^{-1} Z| \\ &= |A_{22}|\,|I - \{I - \delta A_{22} + (\delta A_{22})^2 - \ldots\} Z|. \end{aligned}$$

From this it follows that

$$\log_e |B_{22}| = \log_e |A_{22}| - \operatorname{tr}(Z) + \operatorname{tr}\{(\delta A_{22}) Z\} - \operatorname{tr}\{(\delta A_{22})^2 Z\} - \tfrac{1}{2}\operatorname{tr}(Z^2) + O(\delta^5).$$

We also have

$$\operatorname{tr}(B_{22}) = \operatorname{tr}(A_{22}) - \operatorname{tr}(Z).$$

Hence the expression (1) may finally be written as

$$[-\log_e |A_{22}| + \operatorname{tr}(A_{22}) - q] - \operatorname{tr}\{(\delta A_{22}) Z\} + \operatorname{tr}\{(\delta A_{22})^2 Z\} + \tfrac{1}{2}\operatorname{tr}(Z^2) + O(\delta^5). \qquad (4)$$

The part of (4) within the square brackets is merely what would arise if we were testing the equality of all the latent roots of the covariance matrix A_{22}, of order q. Using Bartlett's result (IIIa) we can evaluate its expectation as

$$\tfrac{1}{2}q(q+1)\left\{\frac{1}{n} + \frac{1}{6n^2}\left(2q + 1 - \frac{2}{q+1}\right)\right\} + O\left(\frac{1}{n^3}\right).$$

To find the expectations of the remaining terms of (4), and of other quantities occurring later, we need the results exemplified as follows, where quantities of order $1/n^3$ are neglected:

$$\begin{aligned} E\{(\delta a_{11})^3\} &= 8\lambda_1^3/n^2, & E\{\delta a_{11}(\delta a_{12})^2\} &= 2\lambda_1^2\lambda_2/n^2, \\ E\{\delta a_{12}\delta a_{13}\delta a_{23}\} &= \lambda_1\lambda_2\lambda_3/n^2, & E\{(\delta a_{11})^4\} &= 12\lambda_1^4/n^2, \\ E\{(\delta a_{11})^2(\delta a_{22})^2\} &= 4\lambda_1^2\lambda_2^2/n^2, & E\{(\delta a_{11})^2(\delta a_{12})^2\} &= 2\lambda_1^3\lambda_2/n^2, \\ E\{(\delta a_{11})^2(\delta a_{23})^2\} &= 2\lambda_1^2\lambda_2\lambda_3/n^2, & E\{(\delta a_{12})^4\} &= 3\lambda_1^2\lambda_2^2/n^2, \\ E\{(\delta a_{12})^2(\delta a_{13})^2\} &= \lambda_1^2\lambda_2\lambda_3/n^2, & E\{(\delta a_{12})^2(\delta a_{34})^2\} &= \lambda_1\lambda_2\lambda_3\lambda_4/n^2. \end{aligned}$$

Terms of other kinds which arise all have expectations of order $1/n^3$ or less.

By making use of the above results, substituting for Z from (2) and neglecting terms of order $1/n^3$ we find that

$$\begin{aligned} E[\operatorname{tr}\{(\delta A_{22}) Z\}] &= \frac{q(q+1)}{n^2}\left\{\sum_{r=1}^{k}\frac{\lambda_r}{\lambda_r - 1} + \sum_{r=1}^{k}\frac{\lambda_r}{(\lambda_r - 1)^2}\right\} \\ &= q(q+1)\,\Sigma_2/n^2, \\ E[\operatorname{tr}\{(\delta A_{22})^2 Z\}] &= q(q+1)\,\Sigma_1/n^2, \\ E[\operatorname{tr}(Z^2)] &= E[\operatorname{tr}\{\delta A_{21}(\Lambda - 1)^{-1}\delta A_{12}\}^2] \\ &= \{q\Sigma_1^2 + q(q+1)\,\Sigma_2\}/n^2, \end{aligned}$$

where

$$\Sigma_1 = \sum_{r=1}^{k}\left(\frac{\lambda_r}{\lambda_r - 1}\right), \quad \Sigma_2 = \sum_{r=1}^{k}\left(\frac{\lambda_r}{\lambda_r - 1}\right)^2.$$

Hence the expectation of expression (4) is

$$\tfrac{1}{2}q(q+1)\left\{\frac{1}{n}+\frac{1}{6n^2}\left(2q+1-\frac{2}{q+1}\right)\right\}+\frac{q}{2n^2}\{\Sigma_1^2+(q+1)(2\Sigma_1-\Sigma_2)\}$$
$$=\tfrac{1}{2}q(q+1)\left[\frac{1}{n}+\frac{1}{6n^2}\left(2q+1-\frac{2}{q+1}\right)+\frac{1}{n^2}\left\{k-\sum_{r=1}^{k}\frac{1}{(\lambda_r-1)^2}+\frac{\Sigma_1^2}{k}\right\}\right].$$

This means that, for general λ, the correct multiplying factor for (1) is

$$n-k-\frac{1}{6}\left(2q+1-\frac{2}{q+1}\right)-\frac{1}{(q+1)}\left\{\sum_{r=1}^{k}\left(\frac{\lambda_r}{\lambda_r-\lambda}\right)\right\}^2+\lambda^2\sum_{r=1}^{k}\frac{1}{(\lambda_r-\lambda)^2}.$$

If practical use were to be made of this result, estimated values l_r would usually have to be substituted for the true latent roots. If $\lambda_1, \lambda_2, \ldots, \lambda_k$ are all large compared with λ then $\lambda^2\Sigma\dfrac{1}{(\lambda_r-\lambda)^2}$ is small, $\Sigma\left(\dfrac{\lambda_r}{\lambda_r-\lambda}\right)$ is approximately equal to k and the multiplying factor becomes approximately

$$n-k-\frac{1}{6}\left(2q+1-\frac{2}{q+1}\right)-\frac{k^2}{q+1}.$$

3. The case where λ is unknown

Now consider the case where the hypothesis is still that the smallest $p-k$ latent roots of A_0 are equal but where the common value λ is unknown. The approximate χ^2 used for testing the hypothesis now has $\frac{1}{2}(p-k-1)(p-k+2)$ degrees of freedom, one less than before, and may be expressed as

$$-\log_e(l_{k+1}l_{k+2}\ldots l_p)+(p-k)\log_e\{(l_{k+1}+l_{k+2}+\ldots+l_p)/(p-k)\}, \tag{5}$$

multiplied by a factor, which when $k=0$ has been shown by Bartlett (his case III c) to be

$$n-\frac{1}{6}\left(2p+1+\frac{2}{p}\right).$$

With the same notation as before and to the same order of approximation (5) may be replaced by

$$-\log_e|B_{22}|+q\log_e\left\{\frac{1}{q}\operatorname{tr}(B_{22})\right\}.$$

Since, putting $\lambda=1$,

$$q\log_e\left\{\frac{1}{q}\operatorname{tr}B_{22}\right\}-q\log_e\left\{\frac{1}{q}\operatorname{tr}A_{22}\right\}$$
$$=q\log_e\left\{1-\frac{1}{q}(\operatorname{tr}Z)\left(1+\frac{1}{q}\operatorname{tr}\delta A_{22}\right)^{-1}\right\}$$
$$=-\operatorname{tr}Z+\frac{1}{q}(\operatorname{tr}\delta A_{22})(\operatorname{tr}Z)-\frac{1}{q^2}(\operatorname{tr}\delta A_{22})^2(\operatorname{tr}Z)-\frac{1}{2q}(\operatorname{tr}Z)^2+O(\delta^5),$$

the expression (5) may be written as

$$\left[-\log_e|A_{22}|+q\log_e\left\{\frac{1}{q}\operatorname{tr}(A_{22})\right\}\right]+[-\operatorname{tr}\{(\delta A_{22})Z\}+\operatorname{tr}\{(\delta A_{22})^2Z\}+\tfrac{1}{2}\operatorname{tr}(Z^2)]$$
$$+\left[\frac{1}{q}(\operatorname{tr}\delta A_{22})(\operatorname{tr}Z)-\frac{1}{q^2}(\operatorname{tr}\delta A_{22})^2(\operatorname{tr}Z)-\frac{1}{2q}(\operatorname{tr}Z)^2\right]+O(\delta^5). \tag{6}$$

The terms within the first pair of square brackets are those which would arise if we were testing the equality of all q latent roots of A_{22}, the common value being unknown. Their expectation is therefore

$$\tfrac{1}{2}(q-1)(q+2)\left\{\frac{1}{n}+\frac{1}{6n^2}\left(2q+1+\frac{2}{q}\right)\right\}+O\left(\frac{1}{n^3}\right).$$

The expectation of the terms in the second pair of square brackets has already been found to be

$$\frac{q}{2n^2}\{\Sigma_1^2+(q+1)(2\Sigma_1-\Sigma_2)\},$$

where Σ_1 and Σ_2 are as defined in the preceding section.

We now find that

$$\begin{aligned}
E[(\operatorname{tr}\delta A_{22})(\operatorname{tr}Z)] &= \frac{2q}{n^2}\left\{\sum_{r=1}^{k}\frac{\lambda_r}{\lambda_r-1}+\sum_{r=1}^{k}\frac{\lambda_r}{(\lambda_r-1)^2}\right\}\\
&= 2q\Sigma_2/n^2,\\
E[(\operatorname{tr}\delta A_{22})^2(\operatorname{tr}Z)] &= 2q\Sigma_1/n^2,\\
E[(\operatorname{tr}Z)^2] &= E[\{\operatorname{tr}\delta A_{21}(\Lambda-1)^{-1}\delta A_{12}\}^2]\\
&= (q^2\Sigma_1^2+2q\Sigma_2)/n^2.
\end{aligned}$$

Hence the expectation of the terms within the third pair of square brackets in (6) is

$$(\Sigma_2-2\Sigma_1-\tfrac{1}{2}q\Sigma_1^2)/n^2,$$

and the expectation of the whole expression (6) is

$$\begin{aligned}
&\tfrac{1}{2}(q-1)(q+2)\left\{\frac{1}{n}+\frac{1}{6n^2}\left(2q+1+\frac{2}{q}\right)\right\}+\tfrac{1}{2}(q-1)(q+2)(2\Sigma_1-\Sigma_2)/n^2\\
&\quad=\tfrac{1}{2}(q-1)(q+2)\left[\frac{1}{n}+\frac{1}{6n^2}\left(2q+1+\frac{2}{q}\right)+\frac{1}{n_2}\left\{k-\sum_{r=1}^{k}\frac{1}{(\lambda_r-1)^2}\right\}\right].
\end{aligned}$$

Thus, for general λ, the correct multiplying factor for (5) is

$$n-k-\frac{1}{6}\left(2q+1+\frac{2}{q}\right)+\lambda^2\sum_{r=1}^{k}\frac{1}{(\lambda_r-\lambda)^2}.$$

In order to make practical use of this, estimates would have to be substituted for the λ_r, and the now unknown λ would have to be estimated as $(l_{k+1}+l_{k+2}+\ldots+l_p)/q$. If $\lambda_1, \lambda_2, \ldots, \lambda_k$ are all large compared with λ, the last term could be omitted and the approximate value of the multiplying factor obtained from that for $k=0$ by substituting $n-k$ for n and $p-k$ for p, as has been suggested by Bartlett.

Strictly speaking what we have shown in this and in the preceding section is that the use of a certain multiplying factor will give the correct expectation for χ^2, with an error of order $1/n^2$. By similar methods it is possible to show that in each case the variance of χ^2 is correctly given, to the same order of approximation; but the algebra is rather laborious and is for this reason omitted. We have not attempted to investigate the higher moments, though it seems quite likely that these also will have an error only of order $1/n^2$. Even so there is a reasonable amount of justification for the use of the above multiplying factors.

In correspondence Prof. M. S. Bartlett has pointed out to me a distinction which is worth making between the cases of λ known and λ unknown. Suppose that in a factor analysis λ represents that part of the variance of each variate which is due to 'error'. If λ is known,

then acceptance of the hypothesis that the residual latent roots are equal to λ implies that all residual variance is the result of error. If, on the other hand, λ is unknown, all that we can infer from the equality of the residual roots is that the residual variance is the result of specific factors and error; we cannot distinguish between these two sources of variation.

4. Parameters of the distribution of the first k latent roots

It is perhaps of some interest to consider briefly the expression (3) for the rth latent root l_r $(r = 1, 2, \ldots, k)$ of A. Taking the expectation we have

$$E(l_r) = \lambda_r + \frac{\lambda_r}{n} \sum_{i=1}^{p}{}' \left(\frac{\lambda_i}{\lambda_r - \lambda_i} \right) + O\left(\frac{1}{n^2} \right),$$

where Σ' is as previously defined. Similarly for the variance of l_r we have

$$\frac{2\lambda_r^2}{n} \left\{ 1 - \frac{1}{n} \sum_{i=1}^{p}{}' \left(\frac{\lambda_i}{\lambda_r - \lambda_i} \right)^2 \right\} + O\left(\frac{1}{n^3} \right);$$

and the third and fourth cumulants of l_r are given by

$$\kappa_3(l_r) = \frac{8\lambda_r^3}{n^2} + O\left(\frac{1}{n^3} \right),$$

$$\kappa_4(l_r) = \frac{48\lambda_r^4}{n^3} + O\left(\frac{1}{n^4} \right).$$

It may also be noted that the covariance between l_r and l_s $(r \neq s)$ is

$$\frac{2}{n^2} \left(\frac{\lambda_r \lambda_s}{\lambda_r - \lambda_s} \right)^2 + O\left(\frac{1}{n^3} \right).$$

A better estimate of λ_r than l_r, having a bias only of order $1/n^2$, is provided by

$$\hat{\lambda}_r = l_r \left\{ 1 - \frac{1}{n} \sum_{i=1}^{k}{}' \left(\frac{l_i}{l_r - l_i} \right) - \frac{p-k}{n} \frac{\lambda}{l_r - \lambda} \right\},$$

where in the summation i does not take the value r and where an estimate is substituted for λ if necessary. The variance of this is found to be

$$\frac{2\lambda_r^2}{n} \left\{ 1 + \frac{1}{n} \sum_{i=1}^{p}{}' \left(\frac{\lambda_i}{\lambda_r - \lambda_i} \right)^2 \right\} + O\left(\frac{1}{n^3} \right).$$

This gives an idea of how near the distribution of $\hat{\lambda}_r$ is to that of a variance estimate with n degrees of freedom.

5. Residual latent roots of a correlation matrix

We consider, finally, the more difficult problem concerning the latent roots of a correlation matrix. To simplify matters we may, without loss of generality, assume that the true variances of the p variates x_i are all unity, so that the true correlation matrix is equal to the covariance matrix A_0 (not now diagonal). Assuming that the first k of the latent roots λ_i of this matrix are distinct, we wish to test the hypothesis that the residual $p-k$ latent roots are equal, the common value λ being, however, unknown.

Let R be the sample correlation matrix and let l_i $(i = 1, 2, \ldots, p)$ represent the latent roots, in order of magnitude, of R. Then the criterion used for testing the above hypothesis (corresponding to Bartlett's case III c) may be written as

$$n[-\log_e(l_{k+1}l_{k+2}\ldots l_p) + (p-k)\log_e\{(l_{k+1}+l_{k+2}+\ldots+l_p)/(p-k)\}], \tag{7}$$

where we have taken the multiplying factor simply as n, since this is sufficiently accurate for our purpose.

The difficulty here is that the criterion does not, even asymptotically, follow a χ^2 distribution, though it will approximately do so if $\lambda_1, \lambda_2, \ldots, \lambda_k$ are large and λ small. Even then, as Bartlett has remarked, the effective number of degrees of freedom depends on the amount of variance removed from each variate by the first k components. Our object is to determine the effective number of degrees of freedom in the general case, in other words to find the expectation of (7). We shall content ourselves with a lower order of approximation than before and shall neglect terms of order $1/n$ or less in this expectation.

Let V and δV be the diagonal matrices whose ith elements are respectively a_{ii} and δa_{ii}. Then

$$V = I + \delta V,$$

and

$$R = V^{-\frac{1}{2}} A V^{-\frac{1}{2}}.$$

Let $Q = [Q_1 : Q_2]$ be an orthogonal matrix whose first k columns, Q_1, represent the column vectors corresponding, in correct order, to the latent roots $\lambda_1, \lambda_2, \ldots, \lambda_k$ of A_0. This means that

$$A_0 Q = Q \begin{bmatrix} \Lambda & 0 \\ 0 & \lambda I \end{bmatrix},$$

where Λ is as previously defined.

Now the latent roots of R are the same as those of AV^{-1}, or of

$$\begin{aligned} Q'AV^{-1}Q &= Q'(A_0 + \delta A)(I + \delta V)^{-1} Q \\ &= Q'A_0 Q + Q'\delta A Q - Q'A_0 \delta V Q + O(\delta^2), \end{aligned}$$

a matrix whose non-diagonal elements are all of order δ. Hence if we neglect the squares of these, the residual latent roots $l_{k+1}, l_{k+2}, \ldots, l_p$ may be equated to the latent roots of the sub-matrix consisting of the last $p-k$ rows and columns, which is

$$H = \lambda I + Q_2'(\delta A - \lambda \delta V) Q_2 + O(\delta^2).$$

The substitution of the determinant and trace of H for the product and sum of the residual latent roots in (7) will involve an error only of order δ^3, so to this order of approximation the expression (7) may be written as

$$n\left\{-\log_e |H| + (p-k)\log_e\left(\frac{1}{p-k}\operatorname{tr} H\right)\right\}.$$

This is equivalent to

$$\begin{aligned} &\frac{n}{2\lambda^2}\left[\operatorname{tr}\{Q_2'(\delta A - \lambda\delta V) Q_2\}^2 - \frac{1}{(p-k)}\{\operatorname{tr} Q_2'(\delta A - \lambda \delta V) Q_2\}^2 + O(\delta^3)\right] \\ &\qquad = \frac{n}{2\lambda^2}\left[\operatorname{tr}\{C(\delta A - \lambda\delta V)\}^2 - \frac{1}{(p-k)}\{\operatorname{tr} C(\delta A - \lambda\delta V)\}^2 + O(\delta^3)\right], \end{aligned} \tag{8}$$

where $C = Q_2 Q_2' = I - Q_1 Q_1'$.

In order to evaluate the expectation of (8) we make use of the well-known result

$$nE(\delta a_{gh}\delta a_{ij}) = \alpha_{gi}\alpha_{hj} + \alpha_{gj}\alpha_{hi},$$

where, as before, α_{ij} denotes the true covariance (or correlation) between x_i and x_j. We then find that

$$\begin{aligned} nE\{\mathrm{tr}\,(C\delta A)^2\} &= (\mathrm{tr}\,CA_0)^2 + \mathrm{tr}\,(CA_0)^2 \\ &= \lambda^2(\mathrm{tr}\,C)^2 + \lambda^2\,\mathrm{tr}\,C \\ &= \lambda^2(p-k)(p-k+1), \\ nE\{(\mathrm{tr}\,C\delta A)^2\} &= 2\,\mathrm{tr}\,(CA_0)^2 \\ &= 2\lambda^2(p-k), \\ nE\{\mathrm{tr}\,(C\delta AC\delta V)\} &= 2\lambda^2\sum_{i=1}^{p} c_{ii}^2, \\ nE\{(\mathrm{tr}\,C\delta A)(\mathrm{tr}\,C\delta V)\} &= 2\lambda^2\sum_i c_{ii}^2, \\ nE\{\mathrm{tr}\,(C\delta V)^2\} &= 2\sum_{i=1}^{p}\sum_{j=1}^{p} c_{ij}^2\alpha_{ij}^2, \\ nE\{(\mathrm{tr}\,C\delta V)^2\} &= 2\sum_i\sum_j c_{ii}c_{jj}\alpha_{ij}^2, \end{aligned}$$

where c_{ij} is the typical element of C.

Using these results and neglecting quantities of order $1/n$, we find the expectation of (8) to be

$$\tfrac{1}{2}(p-k-1)(p-k+2) - \frac{1}{(p-k)}\{2(p-k-1)\lambda\sum_i(c_{ii}^2) - (p-k)\sum_i\sum_j(c_{ij}^2\alpha_{ij}^2) + \sum_i\sum_j(c_{ii}c_{jj}\alpha_{ij}^2)\}. \tag{9}$$

This expression represents the effective number of degrees of freedom for the approximate χ^2. In practice the quantities λ, c_{ij} and the correlation coefficients α_{ij} would usually have to be replaced by sample estimates.

In the limiting case where all correlations tend to unity and $\lambda \to 0$, the number of degrees of freedom becomes $\frac{1}{2}(p-k-1)(p-k+2)$. Furthermore, expression (8) is then asymptotically equivalent to

$$\frac{n}{2\lambda^2}\left[\mathrm{tr}\,\{(Q_2'\delta AQ_2)^2\} - \frac{1}{(p-k)}\{\mathrm{tr}\,(Q_2'\delta AQ_2)\}^2\right],$$

which is of the correct χ^2 form. This is so because the $\frac{1}{2}(p-k)(p-k+1)$ distinct elements of the matrix $Q_2'\delta AQ_2$ are asymptotically normally and independently distributed, the diagonal elements having variance $2\lambda^2/n$ and the non-diagonal elements having variance λ^2/n.

In the particular case considered by Bartlett (1951) all true correlations are equal to ρ and the $p-1$ residual roots of R are to be tested for equality after the effect of the first has been removed. Here the first true latent root, λ_1, is $1+(p-1)\rho$, the residual roots are all $1-\rho$ and the latent column vector Q_1 corresponding to λ_1 is $\frac{1}{\sqrt{p}}\{1, 1, \ldots, 1\}$. Hence putting $\lambda = 1-\rho$, $\alpha_{ii} = 1$, $\alpha_{ij} = \rho$, $c_{ii} = (p-1)/p$, $c_{ij} = -1/p$ $(i \neq j)$ we have

$$\begin{aligned} \lambda\sum_i(c_{ii}^2) &= \frac{(p-1)^2}{p}(1-\rho), \\ \sum_i\sum_j(c_{ij}^2\alpha_{ij}^2) &= \frac{(p-1)}{p}\{(p-1)+\rho^2\}, \\ \sum_i\sum_j(c_{ii}c_{jj}\alpha_{ij}^2) &= \frac{(p-1)^2}{p}\{1+(p-1)\rho^2\}, \end{aligned}$$

and the expression (9) becomes

$$\tfrac{1}{2}(p-2)(p+1)-\frac{(p-1)(p-2)}{p}(1-\rho)^2.$$

This agrees with the value found by Bartlett.

I am indebted to Prof. T. W. Anderson for showing me the draft of a paper by him on 'Asymptotic theory for principal component analysis', which is to be published at a later date. In it he suggests that a better approximation to the distribution of the criterion considered in this section, would be obtained by using $c\chi^2_d$, where c and d are constants and where χ^2_d denotes a χ^2 variate with d degrees of freedom. In order to fit the constants c and d it would, however, be necessary to find the variance as well as the expectation of (7), and this would clearly be an unpleasantly complicated expression.

Note added in proof. The conjecture that use of the correct multiplying factors for the criteria of §§ 2 and 3 would make all moments agree with those of χ^2, neglecting quantities of order $1/n^2$, is now known to be correct. This follows from a general result, to be published later, concerning likelihood ratio criteria.

REFERENCES

BARTLETT, M. S. (1951). The effect of standardization on a χ^2 approximation in factor analysis. *Biometrika*, **38**, 337.

BARTLETT, M. S. (1954). A note on the multiplying factors for various χ^2 approximations. *J. R. Statist. Soc.* B, **16**, 296.

THE VARIANCE OF THE MEAN OF SYSTEMATIC SAMPLES

By R. M. WILLIAMS

Applied Mathematics Laboratory, D.S.I.R., New Zealand

1. Introduction

Systematic sampling methods have been used in a variety of fields, notably in forestry and ecological studies, because of their convenience in the field and their advantages over a randomized scheme if the survey is also to be used for mapping. Their use was justified by work such as that of Hasel (1938) and Osborne (1942), who examined the sampling errors of random, stratified random and systematic surveys, by subjecting detailed forestry survey data to analysis by these three methods. These analyses showed that the systematic designs were generally the most efficient. They left unsolved the problem of estimating the error of the systematic sample, although Osborne's paper gave a method based on computing the mean square serial correlation which gave reasonable agreement with experimental results.

More recently, Finney (1948), also using data from forest surveys, has compared the variances of systematic samples (computed by the overlapping group method to obtain reasonable precision), random samples and stratified random samples with one and two members per stratum. In these cases Finney showed that the variance of the systematic sample differed little from the variance of a stratified random sample of the same size with one observation per stratum, but was appreciably smaller than the variance obtained with half the number of strata and two observations per stratum, which appeared to be the most efficient system giving an unbiased estimate of the variance without supplementary information. Further investigation of the variance of systematic sampling with various models seems likely to be useful.

Yates (1948) examined the sampling variance for a number of particular cases and gave a form (p. 362) which, with supplementary information, provided an estimate of the sampling variance of the means of a systematic sample, or, alternatively, subject to certain assumptions, provided an upper limit to the error. Jowett (1952) gave a method for the determination of the variance of a systematic sample, on the assumption that the observations were derived from a stationary process. This was derived from earlier work by Cochran (1946) and Quenouille (1949) in the same field.

In this paper we shall develop a method similar to Jowett's, but involving rather weaker assumptions; we shall also derive a form for the variance of samples with spacings less than or equal to that of the observed data (which is not covered by Jowett's method); this includes the case where the variance is derived from a single sample. This does not conflict with the often-stated principle that a systematic sample cannot by itself provide an estimate of error, since the assumptions which are made about the population are equivalent to supplementary information. Many alternative assumptions could have been made, and the justification of the particular ones chosen must be that they appear likely to apply to a wide variety of data and in the cases considered below yield results which agree with experiment.

These investigations arose out of discussions with officers of the North Canterbury Catchment Board (New Zealand) on the spacing of points along a line on which observations were to be made to determine the percentage of ground with various types of cover. The main part of the theory is applicable to any series of observations spaced out uniformly in time or space which fit the model described below and holds for both discrete and continuous variables.

We shall, for convenience, always refer to the transect, meaning the whole line along which observations are made, and to the points on it. The details of the experimental work are given in § 4.

2. Formulae for variances

We regard the transect as composed of kn points; the sample is drawn by taking every kth point giving k possible samples. The variable at the ith point is x_i ($i = 1, 2, \ldots, kn$). The formula for the sampling variance $\sigma_s^2(n, k)$ of the mean of a systematic sample (defined as the mean square of the deviation of the sample mean from the population mean $\bar{x}$) drawn from a finite population is given by Madow & Madow (1944) in the first of a series of three papers (Madow, 1949, 1953), and may be derived in a convenient form from the usual partition of sums of squares and the relationship

$$\sum_{i=1}^{kn}(x_i-\bar{x})^2 = \sum_{i=1}^{kn}\sum_{j>i}(x_i-x_j)^2/kn,$$

which gives
$$\sigma_s^2(n,k) = \frac{1}{(kn)^2}\sum_{i=1}^{kn}\sum_{j>i}(x_i-x_j)^2 - \frac{1}{kn^2}\sum_{\delta=1}^{n-1}\sum_{i=1}^{k(n-\delta)}(x_i-x_{i+k\delta})^2$$
$$= \left(\frac{f(0)}{n}+2\bar{f}_n\right)-\left(\frac{f(0)}{kn}+2\bar{f}_{kn}\right), \tag{1}$$

where
$$\left.\begin{aligned} f(0) &= \sum_{i=1}^{kn}(x_i-a)^2/kn, \\ \bar{f}_n &= \frac{1}{n}\sum_{\delta=1}^{n-1}\frac{1}{kn}\sum_{i=1}^{k(n-\delta)}(x_i-a)(x_{i+k\delta}-a), \\ \bar{f}_{kn} &= \frac{1}{kn}\sum_{\gamma=1}^{kn-1}\frac{1}{kn}\sum_{i=1}^{kn-\gamma}(x_i-a)(x_{i+\gamma}-a). \end{aligned}\right\} \tag{2}$$

Here a is any constant.

We let k tend to infinity and define s as i/kn, $x(s)$ as the corresponding value of x_i and $\phi(t)$ as

$$\phi(t) = \int_0^{1-t}(x(s)-a)(x(s+t)-a)\,ds. \tag{3}$$

In the limit
$$\bar{f}_n \to \frac{1}{n}\sum_{\delta=1}^{n-1}\phi\left(\frac{\delta}{n}\right) = \bar{\phi}_n,$$
$$\bar{f}_{kn} \to \int_0^1 \phi(t)\,dt = \bar{\phi},$$
$$f(0) \to \phi(0)$$

and thus
$$\sigma_s^2(n,k) \to \frac{\phi(0)}{n} + 2\bar{\phi}_n - 2\bar{\phi} = \sigma_s^2(n). \tag{4}*$$

* Jowett (1952) gave this formula in a slightly different form, Quenouille (1949) gave the form when n is large as well as k.

An alternative form for $\sigma_s^2(n)$ is obtained from the Euler–Maclaurin theorem from which it follows that if $\phi(t)$ has continuous derivatives up to the $(2q+1)$th order

$$\sum_{\delta=0}^{n} \frac{1}{n}\phi\left(\frac{\delta}{n}\right) = \int_0^1 \phi(t)\,dt + \frac{1}{2n}(\phi(0)+\phi(1)) - \sum_{\nu=1}^{q} \frac{B_{2\nu}}{(2\nu)!}\frac{1}{n^{2\nu}}(\phi^{(2\nu-1)}(0) - \phi^{(2\nu-1)}(1)) + (-1)^{q+1}\frac{1}{n^{2q+2}}\int_0^n P_{2q+1}(x)\,\phi^{(2q+1)}\left(\frac{x}{n}\right)dx,$$

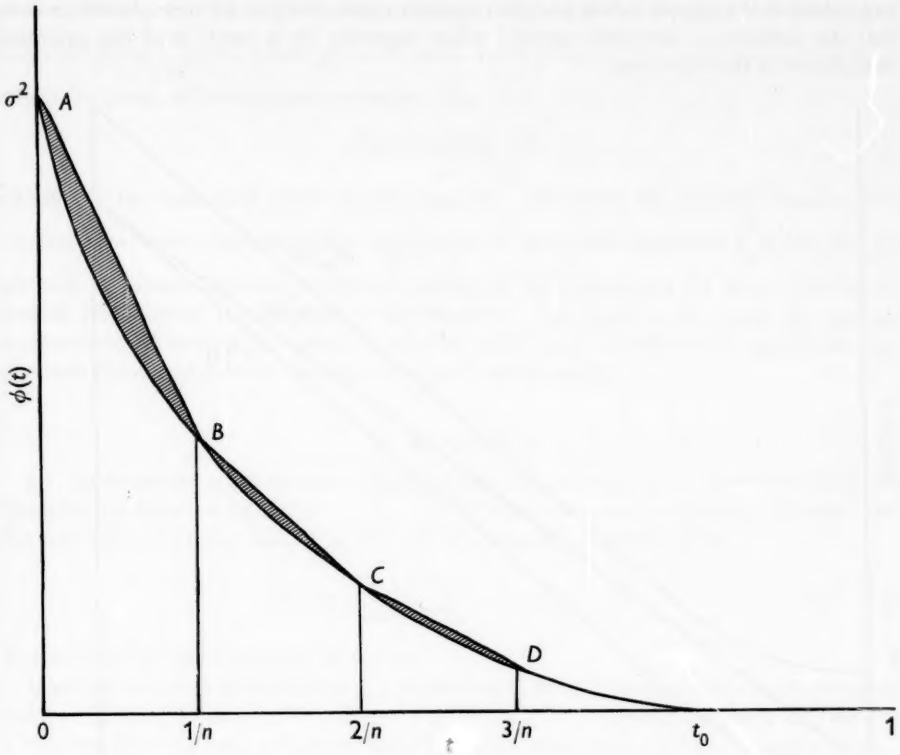


Fig. 1. Graph of $\phi(t)$ showing shaded areas proportional to $\sigma_s^2(n)$. (Not to scale.)

where $\phi^{(i)}(t)$ denotes the ith derivative, $B_{2\nu}$ are the Bernoulli numbers and

$$P_{2q+1}(x) = \sum_{l=1}^{\infty} \frac{\sin(2l\pi x)}{2^{2q}(l\pi)^{2q+1}}.$$

Assuming that the series converges we have

$$\sigma_s^2(n) = \frac{\phi^{(1)}(1)-\phi^{(1)}(0)}{6n^2} - \frac{\phi^{(3)}(1)-\phi^{(3)}(0)}{360n^4} + \frac{\phi^{(5)}(1)-\phi^{(5)}(0)}{15120n^6} + \ldots. \tag{5}$$

The assumption that $\phi(t)$ is differentiable to the required order is one that in general will not be fulfilled by the observed points in a particular transect. If, however, we follow Cochran (1946) and regard the observation of a particular transect as a sample drawn from

a multivariate parent population, then it will often be reasonable to assume that the expected values of $\phi(t)$, etc. (the expectation being taken over all possible drawings from the parent population) do have the required property of differentiability even if the individual values are discontinuous. We shall therefore, in future, consider the expected value of the variance, etc. Since the formulae are unchanged if we replace $\sigma_s^2(n)$, $\phi(t)$, etc., by their expected values, and there is no danger of confusion, we shall use the same symbols for the expected values. Where we wish to emphasize the distinction we shall refer to the two values as the sample values and the expected values of $\sigma_s^2(n)$, $\phi(t)$, etc. It will be seen that the distinction, although essential where equation (5) is used, is of less practical importance in the other case.

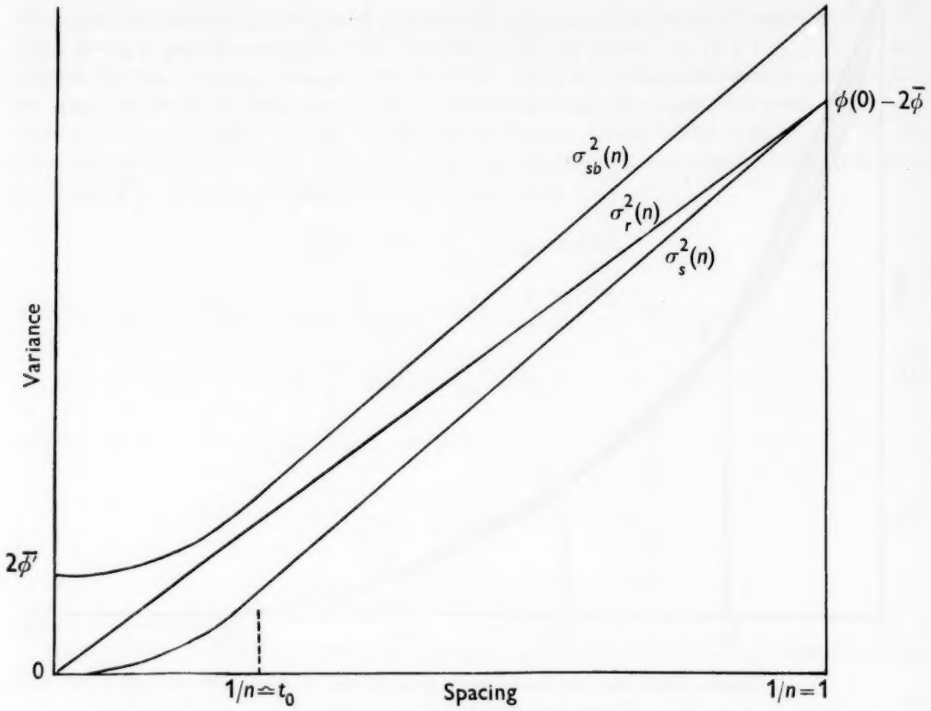


Fig. 2. Relative values of variances for monotonic decreasing $\phi(t)$. (Not to scale.)

By similar methods we find that the sampling variance of the mean of a random sample of n members ($\sigma_r^2(n)$) is given by

$$\sigma_r^2(n) = \frac{1}{n}(\phi(0) - 2\overline{\phi}). \tag{6}$$

The functions involved in (4) and (6) are shown in Fig. 1 for a monotonic decreasing $\phi(t)$; $\overline{\phi}$ is given by the area under the curve $\phi(t)$; $\phi(0)/2n + \overline{\phi}_n$ is the area of the polygon $OABCD\ldots$, so that $\sigma_s^2(n)$ is equal to twice the area shown shaded.*

* This graphical method is equivalent to that given by Jowett (1952). It should however be noted that Jowett's assumption that the x's form a stationary time series is not invoked.

From (5) it follows that provided the expansion is valid $\sigma_s^2(n)$ initially varies as the square of the distance $(1/n)$ between observations; on the other hand, $\sigma_r^2(n)$ varies linearly with $1/n$. For sufficiently large n, $\sigma_r^2(n)$ will therefore be greater than $\sigma_s^2(n)$, and the systematic sample more efficient than the random.

If $\phi(t)$ is zero or changes linearly beyond some point t_0, then as the spacing of the sampling points $(1/n)$ increases beyond t_0, $\sigma_s^2(n)$ will increase linearly with $1/n$ (see Fig. 2).

The variance $\sigma_s^2(n)$ is a measure of the variability of the mean of all samples drawn from the same transect; if we denote by $\sigma_{sb}^2(n)$ the mean square of the deviations of the mean of a sample about $\int_0^1 \mu(t)\,dt$, where $\mu(t)$ is the mean at the point t of the parent population from which the points of the transect are drawn, then

$$\sigma_{sb}^2(n) = \sigma_s^2(n) + 2\overline{\phi}', \tag{7}$$

where $\overline{\phi}'$ is the value of $\overline{\phi}$ when a is put equal to $\int_0^1 \mu(t)\,dt$ (see Fig. 2). This variance is of interest if we wish to determine the significance of suspected changes in $\int_0^1 \mu(t)\,dt$, and we are able to assume that the covariance matrix of the x's remains the same. This would require justification; for example, in the transects considered in this paper, the gradual replacement of one species of plant by another might lead not only to changes in the proportions of covered ground but also in the covariance matrix.

3. ESTIMATION

3·1. To determine $\sigma_s^2(n)$ we must, ideally, know $\phi(t)$ at all points in the range $(0, 1)$. We estimate this from a sample $x_r\,(r = 1, 2, \ldots, m)$ of m points spaced uniformly $1/m$ apart with the first point randomly located in $(0, 1/m)$. For such a sample we define

$$\psi(u/m) = \frac{1}{m-u} \sum_{r=1}^{m-u} (x_r - a)(x_{r+u} - a); \tag{8}$$

$\psi(u/m)$ is an unbiased estimate of $\phi(u/m)$.

If we can assume a specific form (e.g. exponential) for $\phi(t)$ we can fit this to the observed values of $\psi(u/m)$ preferably for values of u sufficiently low to ensure that they are based on a large number of points, and determine $\sigma_s^2(n)$ either graphically or from the series form.

Where this is not possible the determination of $\phi(t)$ presents no difficulties for low values of t and consequently the differential coefficients at $t = 0$, required for the series form, can easily be determined; but as t approaches unity the number of terms involved will be too small to provide satisfactory estimates. For the graphical form the same objection holds, and in addition it is desirable to avoid the labour of computing $\phi(t)$ at all points if possible.

If $\phi(t)$ is linear above some value t_0 then in the graphical method no further contribution to $\sigma_s^2(n)$ will be made by continuing $\phi(t)$ beyond t_0. In practice it will probably be sufficient to continue $\phi(t)$ to a point such that it is effectively linear over lengths equal to the greatest sampling intervals likely to be of interest.

Assuming the expansion is valid, the error involved in considering only the contribution up to t_0 is given by

$$\sum_{\nu=1}^{\infty} \frac{B_{2\nu}}{(2\nu)!\, n^{2\nu}} [\phi^{(2\nu-1)}(1) - \phi^{(2\nu-1)}(t_0)].$$

If the series form is used some equivalent assumption is necessary. If it is assumed that $\phi(t)$ is linear from $t = t_0$ onwards, then $\phi'(1) = -\phi(t_0)/(1-t_0)$ and all higher derivatives are zero. An alternative to this is to assume that in the region $(0, 1-t_0)$, $x(s)$ has a mean μ_1 and is independent of $x(s)$ in the region $(t_0, 1)$, where the mean is μ_2; then for $t > t_0 > \frac{1}{2}$

$$\phi(t) = (1-t)(\mu_1 - a)(\mu_2 - a).$$

Visual inspection of the observations should provide some check on the validity of this assumption, and $\phi(t)$ can then be determined directly from the means at the beginning and end of the transect. Even if this assumption does not hold exactly it does not appear to be very sensitive to deviations. For example, if μ_1 and μ_2 are not constant but each is subject to a different linear trend of the form

$$\mu_1 = \alpha_1 + \beta_1 t, \quad \mu_2 = \alpha_2 + \beta_2 t,$$

then

$$\phi(t) = (1-t)(\alpha_1 - a)(\alpha_2 + \beta_2 - a) + \frac{(1-t)^2}{2}[\beta_1(\alpha_2 + \beta_2 - a) - \beta_2(\alpha_1 - a)] - \frac{(1-t)^3}{6}\beta_1\beta_2,$$

which gives

$$\phi^{(1)}(1) = -(\alpha_1 - a)(\alpha_2 + \beta_2 - a),$$
$$\phi^{(3)}(1) = \beta_1\beta_2.$$

The assumption of constant means in the regions $(0, 1-t_0)$ and $(t_0, 1)$ gives

$$\phi^{(1)}(1) = -(\alpha_1 - a)(\alpha_2 + \beta_2 - a) - \frac{(1-t_0)}{2}[\beta_1(\alpha_2 + \beta_2 - a) - \beta_2(\alpha_1 - a)] + \frac{(1-t_0)^2}{4}\beta_1\beta_2,$$

so that, if t_0 is taken as large as possible consistent with basing the means on a reasonable sample, the error will not be large unless the trend is marked, i.e. β_1 and β_2 are large.

3·2. When a pilot survey (of m points) has been undertaken to provide design data we may want to calculate $\sigma_s^2(n)$ for values of n greater or less than m; where the variance of the mean of a sample is to be calculated from the sample itself $m = n$.

When n is substantially smaller than m, say half or less, the graphical method is best since the series form will converge slowly. When n is either of the same order as or greater than m the sampling fluctuations in determining the values of $\phi(t)$ will be comparable with the differences to be measured, and the series form must be used.

The differential coefficients required at $t = 0$ may be determined by the Gregory–Newton formulae, i.e. by putting

$$\phi^{(1)}(0) = m(\Delta\psi(0) - \tfrac{1}{2}\Delta^2\psi(0) \dots),$$
$$\phi^{(3)}(0) = m^3(\Delta^3\psi(0) - \tfrac{3}{2}\Delta^4\psi(0) \dots),$$

where

$$\Delta\psi(0) = \psi\left(\frac{1}{m}\right) - \psi(0),$$

$$\Delta^2\psi(0) = \psi\left(\frac{2}{m}\right) - 2\psi\left(\frac{1}{m}\right) + \psi(0).$$

Since the sampling variance of $\Delta^i\psi(0)$ increases rapidly with i it will be better to use only a few terms in estimating $\phi^{(1)}(0)$, $\phi^{(3)}(0)$, etc., and for the same reason to use only a few terms in the actual expansion of $\sigma_s^2(n)$. This means that the series expansion is only useful when n is large enough to ensure rapid convergence, since otherwise serious bias may occur. In many practical cases precise estimates of the variance are not needed, and fairly large errors may be tolerated.

3·3. When the observations are subject to a periodic effect it may cause serious errors if the period is close to $1/m$ or a submultiple of it, since in the graphical method it will not be justifiable to draw a smooth curve through the calculated values of $\psi(u/m)$; the series form may not converge sufficiently rapidly to provide reliable results. If the existence of a period can be expected from the nature of the data, then this difficulty can be avoided by making $1/m$ a sufficiently small fraction of the period. Finney (1950) gave an example of an unexpected periodic variation occurring in a forestry survey, for which there appeared to be no simple explanation; the only safeguard against effects of this type is to take the observations so close that it is at least unlikely that the periodic effect will be undetected.

3·4. In the particular case where $x(s)$ has the value 0 or 1, equation (8) has a very simple form if we put $a = 0$. Writing ν_u for the number of 1's followed u places later by another 1 we have

$$\psi\left(\frac{u}{m}\right) = \frac{\nu_u}{m}.$$

In this case we can set limits on the gradient at $t = 1$ since

$$0 \leqslant \phi(t) \leqslant (1-t),$$

so that

$$-1 \leqslant \phi^{(1)}(1) \leqslant 0.$$

If it is assumed that no correlation exists between points at the extreme ends of the transect then $\phi^{(1)}(1) = -p_1 p_2$, where p_1 and p_2 are the proportions of 1's at the beginning and end of the transect.

If m is large so that we can neglect all but the first differences, we have

$$\sigma_s^2(n) \simeq \frac{m_r}{12n^2}, \tag{9}$$

where m_r is the number of runs of 1's and 0's in the set of m observations.

This approximation is only valid if m is so large that, because of the high correlation between adjacent points, the number of runs is altered relatively little by small changes in m. We can, for example, find the number of runs in the series obtained by omitting alternate observations; if this differs little from m_r we can use this form as a quick estimate of the variance.

4. Experimental results

The experimental data consisted of a record of vegetative cover which was classified as either living vegetation, dead vegetation or bare ground. The vegetation comprised a portion of the native tussock grasslands that occur on the mountainous area eastwards of the main divide of the Southern Alps of the South Island, New Zealand. Field work was carried out in the Waimakariri river catchment, where the soil is largely derived from steep and rugged mountains, composed of hard greywacke rock, which rise to some 7000 ft. above sea-level. For a century sheep have grazed the tussock grasslands which have been divided by natural barriers and fences into extensive blocks.

The point method of pasture analysis was used and line transects of 10 chains and more in length were permanently pegged. The point analyser consisted of a metal frame 20 in. long which carried a row of steel pins 2 in. apart. At the same calendar date in the summer of each year the analyser was placed at a series of adjacent positions along the transect and

Table 1. *The number v_u of points followed by a point with similar cover at spacing u*

u	Transect 4			Transect 7			Transect 11			Transect 8
	L	*D*	*B*	*L*	*D*	*B*	*L*	*D*	*B*	*L*
0	1570	201	2226	4713	343	1227	2955	127	676	3335
1	1362	58	1992	4315	84	933	2787	33	514	2750
2	1252	38	1895	4194	55	848	2718	24	456	2498
3	1169	29	1815	4116	43	802	2668	14	424	2340
4	1100	25	1745	4061	36	750	2635	16	401	2214
5	1046	27	1681	4015	31	717	2601	12	381	2151
6	974	20	1623	3991	33	694	2574	11	365	2077
7	919	16	1579	3974	27	668	2556	8	350	2042
8	872	19	1530	3936	21	647	2534	6	338	2025
9	829	13	1488	3924	24	630	2518	8	320	1999
10	797	11	1454	3909	18	626	2506	4	314	1974
11	773	9	1426	3888	18	615	2499	3	306	1962
12	762	11	1411	3869	18	595	2478	4	294	1968
13	743	15	1398	3855	17	586	2464	7	281	1974
14	733	11	1382	3856	21	581	2451	6	271	1981
15	724	8	1369	3847	17	581	2444	6	268	1979
16	724	9	1369	3832	18	561	2434	8	260	1966
17	730	4	1373	3812	18	550	2427	7	259	1936
18	728	12	1368	3809	17	557	2418	6	251	1927
19	719	16	1363	3788	14	544	2414	3	250	1915
20	718	11	1363	3784	13	530	2408	4	245	1936
21	709	14	1348	3786	21	535	2404	1	243	1951
22	704	13	1356	3788	24	522	2407	5	253	1941
23	693	11	1330	3779	17	524	2404	5	249	1945
24	694	11	1325	3778	21	522	2389	8	236	1942
25	696	10	1322	3767	25	513	2380	3	232	1944
26	700	10	1326	3755	22	501	2387	5	236	1958
27	703	14	1335	3745	16	492	2395	4	236	1945
28	—	—	—	3745	19	488	2398	4	240	1931
29	—	—	—	3746	14	483	2396	4	244	1918
30	—	—	—	3745	19	473	2398	9	237	1910
35	—	—	—	—	—	—	—	—	—	1875
40	—	—	—	—	—	—	—	—	—	1826
No. of points (m)	3997			6283			3758			6321
Percentage of cover	39·28	5·03	55·69	75·01	5·46	19·53	78·63	3·38	17·99	52·76

the state of the cover beneath each pin was recorded; this gave a set of observations uniformly spaced every 2 in. along the transect.

Four transects covering a wide range of conditions were chosen for study; these are referred to as transects 4, 7, 8 and 11. A brief description of each is given in the Appendix.

We consider separately the three cases living or not living, dead or not dead, bare or not bare; the first is the most important ecologically. We describe the state of the ground at the point s by putting $x(s) = 1$ if there is living cover, $x(s) = 0$ if not, and similarly for the other cases.

Table 2. *The series for* $\sigma_s^2(n)$

Transect 4, $m = 3997$

$$V_L = \frac{10^4}{n^2}\left(46{\cdot}75 - 0{\cdot}197\,\frac{m^2}{n^2} \ldots\right)$$

$$V_D = \frac{10^4}{n^2}\left(40{\cdot}30 - 0{\cdot}311\,\frac{m^2}{n^2} \ldots\right)$$

$$V_B = \frac{10^4}{n^2}\left(57{\cdot}03 - 0{\cdot}381\,\frac{m^2}{n^2} \ldots\right)$$

Transect 7, $m = 6283$

$$V_L = \frac{10^4}{n^2}\left(102{\cdot}32 - 0{\cdot}650\,\frac{m^2}{n^2} \ldots\right)$$

$$V_D = \frac{10^4}{n^2}\left(74{\cdot}17 - 0{\cdot}592\,\frac{m^2}{n^2} \ldots\right)$$

$$V_B = \frac{10^4}{n^2}\left(75{\cdot}85 - 0{\cdot}473\,\frac{m^2}{n^2} \ldots\right)$$

Transect 11, $m = 3758$

$$V_L = \frac{10^4}{n^2}\left(40{\cdot}59 - 0{\cdot}222\,\frac{m^2}{n^2} \ldots\right)$$

$$V_D = \frac{10^4}{n^2}\left(27{\cdot}53 - 0{\cdot}239\,\frac{m^2}{n^2} \ldots\right)$$

$$V_B = \frac{10^4}{n^2}\left(39{\cdot}99 - 0{\cdot}217\,\frac{m^2}{n^2} \ldots\right)$$

Transect 8, $m = 6321$

$$V_L = \frac{10^4}{n^2}\left(138{\cdot}48 - 0{\cdot}664\,\frac{m^2}{n^2} \ldots\right)$$

In Table 1 we give the values of ν_u for living, dead and bare cover (L, D, B) for the cases considered.

Examination of Table 1 shows that for the observations of dead points for spacings greater than about 50 in. (i.e. $u > 25$) the value of ν_u (apart from random fluctuations) is approximately that which would be expected if there was no correlation (i.e. $(m-u)p^2$, where p is the proportion dead). For the living and bare points there is still some correlation at this spacing, but it is changing sufficiently slowly to regard it as linear for all except large sampling intervals; this is equivalent to ignoring any contribution to the variance made by values of $\phi(t)$ not tabulated.

To obtain the coefficients in the series expansion only differences up to the third were used at $t = 0$ in order to avoid excessive sampling errors. At $t = 1$, $\phi^{(1)}(1)$ was taken as $-p^2$; all other derivatives were taken as zero.

In Table 2 we give the variances of the percentages of living, dead and bare (V_L, V_D, V_B) in series form. The total number of points in the transect is m so that n sampling points correspond to a spacing of $2m/n$ inches.

Table 3. *Comparison between the variances of systematic and random samples drawn from correlated data*

Spacing ($2m/n$ in.)	10 in.	20 in.	40 in.	60 in.	80 in.
			Transect 4		
V_L	0·604 (0·543*)	1·746	6·088 (5·852†)	11·156	16·223 (9·223†)
V'_L	2·996	5·992	11·985	17·977	23·970
V_D	0·370 (0·373*)	0·902	2·091 (3·233†)	3·330	4·568 (2·538†)
V'_D	0·571	1·143	2·286	3·429	4·572
V_B	0·586 (0·610*)	1·959	6·219 (2·059†)	11·118	16·017 (19·519†)
V'_B	3·090	6·180	12·360	18·540	24·720
			Transect 7		
V_L	0·400 (0·420*)	1·141	2·942 (6·200†)	4·991	7·020 (9·120†)
V'_L	1·496	2·992	5·983	8·975	11·967
V_D	0·267 (0·273*)	0·644	1·462 (1·863†)	2·282	3·100 (1·918†)
V'_D	0·409	0·819	1·637	2·456	3·274
V_B	0·341 (0·310*)	0·842	2·331 (3·229†)	3·480	5·300 (7·065†)
V'_B	1·257	2·514	5·029	7·543	10·058
			Transect 11		
V_L	0·435 (0·510*)	1·606	4·090 (4·494†)	7·819	11·549 (15·366†)
V'_L	2·235	4·470	8·940	13·410	17·881
V_D	0·243 (0·277*)	0·633	1·503 (1·769†)	2·444	3·386 (4·796†)
V'_D	0·434	0·869	1·738	2·606	3·475
V_B	0·437 (0·463*)	1·389	3·463 (3·426†)	6·400	9·337 (13·535†)
V'_B	1·917	3·833	7·667	11·500	15·334
			Transect 8		
V_L	0·661 (0·610*)	1·791	4·957 (3·176†)	8·474	11·410 (6·744†)
V'_L	1·969	3·939	7·878	11·817	15·755

* Approximate estimates determined from $\sigma_s^2(n) = m_r/12n^2$.
† Estimates obtained from sampling experiment.

The values of the variance obtained by direct integration are given in Table 3, and for comparison the variances V'_L, V'_D and V'_B which would be obtained with the same sampling number if the observations were made at random. At close spacings V_L and V_B are much less than V'_L and V'_B; the difference is smaller between V_D and V'_D due to the fact that the dead vegetation comprised a comparatively small percentage of the whole and occurred more or less randomly interspersed with the living cover, while the living vegetation, particularly

in transect 4, tended to occur in clumps interspersed with patches of rock or bare ground, leading to high values of $\phi(t)$ at close spacings of the order of 10–20 in. The smaller differences on transect 8 are consistent with the rather more uniform cover.

The values of the variance calculated from the approximate formula given in (9) for a spacing of 10 in. ($n = m/5$) are also shown in Table 3; the agreement is satisfactory. This example illustrates a point noted in § 3·2. The approximate formula is very nearly equivalent to using only $\Delta\psi(0)$ to estimate $\phi^{(1)}(0)$ and ignoring all later terms in the series for $\sigma_s^2(n)$; for a spacing of 10 in. this gives a better approximation to the variance as determined by numerical integration than does the series form given in Table 2, which involves $\Delta^2\psi(0)$ and $\Delta^3\psi(0)$ as well.

As a check on the validity of the theory and indirectly on the adequacy of the model we took 10 samples, starting each at a point randomly chosen from the first 20 points on the transect and taking every 20th successive point and a similar set of 20 samples taking every 40th successive point. The variances of the means of these samples provide an experimental estimate of $\sigma_s^2(n,k)$ given by (1), where $k = 20$ or $k = 40$, and $kn = m$. But $\sigma_s^2(n,k)$ can also be regarded as an estimate of $\sigma_s^2(n) - \sigma_s^2(kn)$, so that by adding the small term $\sigma_s^2(kn)$, given by putting $n = m$ in (9) to the experimental estimate of $\sigma_s^2(n,k)$, we obtain an estimate of $\sigma_s^2(n)$. These are given in brackets next to the corresponding variances determined by the graphical method (spacings 40 and 80 in.) in Table 3. These comparisons are not independent, but the mean of the ratio of the experimental variances to the variances obtained by the graphical method is 1·0946 which is satisfactory.

5. SUMMARY

The formula given by Madow & Madow (1944) for the variance of the mean of a systematic sample drawn from a finite population was applied to a continuous population, and a method was developed of determining, subject to certain assumptions, the variance of the mean of a systematic sample from single samples. This was applied to the problem of estimating the amount of vegetative cover in ecological studies.

I wish to express my thanks to Dr P. Whittle and other members of the Applied Mathematics Laboratory for valuable suggestions; to Mr R. D. Dick, Soil Conservation Research Officer of the North Canterbury Catchment Board, who raised this problem with the laboratory, made available his extensive records, gave much time in discussions and supplied the descriptive notes; also to the referees for bringing certain recent work to my notice.

REFERENCES

COCHRAN, W. G. (1946). Relative accuracy of systematic and stratified random samples for a certain class of populations. *Ann. Math. Statist.* **17**, 164–77.

FINNEY, D. J. (1948). Random and systematic sampling in timber surveys. *Forestry*, **22**, 64–99.

FINNEY, D. J. (1950). An example of periodic variation in forest sampling. *Forestry*, **23**, 96–111.

HASEL, A. A. (1938). Sampling error in timber surveys. *J. Agric. Res.* **57**, 713–36.

JOWETT, G. H. (1952). The accuracy of systematic sampling from conveyor belts. *Appl. Statist.* **1**, 50–9.

MADOW, W. G. (1949). On the theory of systematic sampling. II. *Ann. Math. Statist.* **20**, 333–54.

MADOW, W. G. (1953). On the theory of systematic sampling. III. *Ann. Math. Statist.* **24**, 101–6.

MADOW, W. G. & MADOW, L. H. (1944). On the theory of systematic sampling. I. *Ann. Math. Statist.* **15**, 1–24.

OSBORNE, J. G. (1942). Sampling errors of systematic and random surveys of cover-type areas. *J. Amer. Statist. Ass.* **37**, 256–64.
QUENOUILLE, M. H. (1949). Problems in plane sampling. *Ann. Math. Statist.* **20**, 355–75.
YATES, F. (1948). Systematic sampling. *Phil. Trans.* A, **241**, 345–77.

APPENDIX: DESCRIPTIVE NOTES

Transect 4, *Leith Hill*

The top peg is at 4250 ft. and the line transect extends down the mountain slope for 10 chains at an average angle of 29°. In the upper portion, particularly, islands of vegetation (tall tussocks, *Danthonia flavescens*) are surrounded by bare ground with the top soil eroded, and there remains the yellowish subsoil down to little subsoil amongst rock fragments. On the lower slopes the tussocks *Festuca novae zealandae* and *Poa colensoi* form the tussock content of the grassland. Over two dozen species of native and introduced grasses and herbaceous plants grow between the tussocks as well as a few shrubby plants, such as *Pimelia prostrata* and a prostrate *Dracophyllum* species.

Transect 7, *Bridge Hill*

The top peg is at 3450 ft., and the line transect descends at an average angle of 33° for a distance of 10 chains. The perennial herb *Celmisia spectabilis* with the tussock *Festuca novae zealandae* and the prostrate much-branched shrub *Gaultheria depressa*, dominate the physiognomic features in this portion of the tussock grasslands.

Transect 8, *Constitution Hill*

The top peg is at 4000 ft., and the line transect descends at an average angle of 33° for a distance of 15 chains. *Festuca novae zealandae* and *Poa colensoi* are the two common tussock species, and about two dozen native and introduced species of grasses and herbs grow in the inter-tussock spaces. Shrubby species, such as *Pimelia prostrata, Cassinia fulvida, Discaria toumatou* and *Leptospermum scoparium* occur occasionally.

Transect 11, *Lower Lyndon*

The top peg is at 3550 ft. and the line transect descends at an average angle of 29°. This portion of the tussock grasslands is characterized by shrubby growth comprising *Cyathodes colensoi, Gaultheria* species, *Discaria toumatou* interspersed with *Celmisia spectabilis, Festuca novae zealandae* and representatives of about the two dozen species of native and introduced grasses and herbs common in these grasslands.

GROUPING METHODS IN THE FITTING OF POLYNOMIALS TO UNEQUALLY SPACED OBSERVATIONS

By P. G. GUEST

University of Sydney, Australia

Introduction

When a polynomial is to be fitted to a series of observations y_i at values x_i of the independent variable and the number of observations n is large, the observations at neighbouring values of x_i are often grouped together to give a set of $N = n/r$ pairs of values y_{Ni}, x_{Ni}. These quantities y_{Ni}, x_{Ni}, are the sums of the r observed values in the particular group. Since the labour required to fit a polynomial curve increases very greatly as the number of observations becomes larger, the time spent in fitting a curve to the N grouped observations is only a small fraction of that which would be required if the original set of n observations was used. It is for this reason that grouping is often employed before fitting a curve to the observed values.

It is clear, however, that in general the grouped curve will not be the same as the curve obtained from the original set of observations. The random sampling errors will affect the curves differently in the two cases, but that is a minor matter. What is more important is that the grouping will give rise to a loss of information—an increase in standard error and a decrease in efficiency—and may also give rise to bias in the estimates. It is the purpose of the present paper to investigate the last two effects and to develop criteria by means of which the suitability of the grouping method as an approximation to the full least-squares solution may be determined in any particular case.

In an earlier paper (Guest, 1954) an account has been given of the use of grouping methods in the fitting of polynomials to equally spaced observations. Unfortunately, no exact general treatment can be given in cases where the observations are not spaced at equal intervals. The procedure adopted here is to describe the departure from uniform spacing by two parameters κ_2 and κ_3. κ_2 is a measure of the crowding of the observations towards one end of the range of the independent variable x, while κ_3 is a measure of the crowding towards the centre of the range rather than towards the ends. A full account of the use of these parameters has been given in a previous paper (Guest, 1953*a*).

A short section has been included on the use of step functions in the fitting of polynomials, with a full discussion of the cases where the polynomial is only of the first degree. Finally, the various methods of obtaining a fitted curve are compared from the point of view of time required, of efficiency, and of bias of the estimates.

Representation of the observations

When the values of the independent variable x_i at the n points of observation are arranged in order of magnitude, each observation may be identified by the number ϵ giving its position in the sequence, ϵ taking integral or half-integral values from $+\frac{1}{2}(n-1)$ to $-\frac{1}{2}(n-1)$. The observations at the points $x_n(\epsilon)$ will be represented by the symbols $y_n(\epsilon)$. In the present discussion the system of points $x_n(\epsilon)$ will be replaced by a smoothed-out system $X_n(\epsilon)$

obtained by fitting a curve of the third degree in ϵ to the values $x_n(\epsilon)$. With a suitable choice of origin

$$X_n(\epsilon) = k_{1n}T_{1n}(\epsilon) + k_{2n}T_{2n}(\epsilon) + k_{3n}T_{3n}(\epsilon), \tag{1}$$

where

$$k_{jn} = \sum_{\epsilon} x_n(\epsilon)\, T_{jn}(\epsilon) / \sum_{\epsilon} T_{jn}^2(\epsilon), \tag{2}$$

and $T_{jn}(\epsilon)$ is the orthogonal polynomial of degree j in ϵ.

If the n observations are converted by grouping into $N = n/r$ values

$$y_N(\epsilon) = \sum_{z=-\frac{1}{2}(r-1)}^{\frac{1}{2}(r-1)} y_n(r\epsilon + z)$$

at points

$$x_N(\epsilon) = \sum_{z} x_n(r\epsilon + z),$$

the smoothed-out system for the grouped observations may be written as

$$X_N(\epsilon) = k_{1N}T_{1N}(\epsilon) + k_{2N}T_{2N}(\epsilon) + k_{3N}T_{3N}(\epsilon), \tag{3}$$

where from the equations connecting the polynomials $T_{jn}(\epsilon)$, $T_{jN}(\epsilon)$ (Guest, 1954)

$$\left.\begin{aligned} k_{1N} &= r^2[k_{1n} - \tfrac{1}{10}(r^2 - 1)\, k_{3n}], \\ k_{2N} &= r^3 k_{2n}, \\ k_{3N} &= r^4 k_{3n}. \end{aligned}\right\} \tag{4}$$

It is convenient to remove a scale factor

$$\phi_n = (k_{1n} + \tfrac{1}{10}n^2 k_{3n})^{-1} \tag{5}$$

from equation (1), and to change to polynomials

$$\tau_{jn}(\epsilon) = n^{-j+1}T_{jn}(\epsilon). \tag{6}$$

Then

$$X_n(\epsilon) = \phi_n^{-1}\xi_n(\epsilon), \tag{7}$$

where

$$\xi_n(\epsilon) = \kappa_{1n}\tau_{1n}(\epsilon) + \kappa_{2n}\tau_{2n}(\epsilon) + 2\kappa_{3n}\tau_{3n}(\epsilon), \tag{8}$$

$$\kappa_{1n} = \phi_n k_{1n}, \quad \kappa_{2n} = n\phi_n k_{2n}, \quad \kappa_{3n} = \tfrac{1}{2}n^2\phi_n k_{3n}, \tag{9}$$

and

$$\kappa_{1n} = 1 - \tfrac{1}{5}\kappa_{3n}. \tag{10}$$

The advantage of this notation is that the coefficients κ_{jn} are all of the same order of magnitude, as also are the orthogonal polynomials $\tau_{jn}(\epsilon)$ (Guest, 1953*a*).

For the grouped observations $X_N(\epsilon)$ is replaced by

$$\xi_N(\epsilon) = \kappa_{1N}\tau_{1N}(\epsilon) + \kappa_{2N}\tau_{2N}(\epsilon) + 2\kappa_{3N}\tau_{3N}(\epsilon), \tag{11}$$

where

$$X_N(\epsilon) = r^2\phi_n^{-1}\xi_N(\epsilon), \tag{12}$$

and

$$\kappa_{1N} = r^{-2}\phi_n k_{1N}, \quad \kappa_{2N} = Nr^{-2}\phi_n k_{2N}, \quad \kappa_{3N} = \tfrac{1}{2}N^2 r^{-2}\phi_n k_{3N}. \tag{13}$$

From equations (4), connecting k_{jn}, k_{jN}, and equations (9) and (13), it is found that

$$\left.\begin{aligned} \kappa_{3N} &= \kappa_{3n}, \quad \kappa_{2N} = \kappa_{2n}, \\ \kappa_{1N} &= \kappa_{1n} - \tfrac{1}{5}(N^{-2} - n^{-2})\kappa_{3n}. \end{aligned}\right\} \tag{14}$$

It will be assumed that n is so large that n^{-2} can be neglected. Then $\xi_N(\epsilon)$ can be written in the form

$$\xi_N(\epsilon) = \sum_{l=1}^{3} (c_{1l} - N^{-2} d_{1l})\, \tau_{lN}(\epsilon), \tag{15}$$

where

$$\left.\begin{aligned} c_{11} &= 1 - \tfrac{1}{5}\kappa_3, & d_{11} &= \tfrac{1}{5}\kappa_3, \\ c_{12} &= \kappa_2, & d_{12} &= 0, \\ c_{13} &= 2\kappa_3, & d_{13} &= 0, \end{aligned}\right\} \tag{16}$$

and κ_j is written for κ_{jn}.

EFFICIENCIES OF THE ESTIMATES

The fitted curve of degree p in x is usually expressed in the form of a power series,

$$u_p(x) = \sum_{j=0}^{p} b_{pj} x^j. \tag{17}$$

Less commonly, the curve may be expressed as a series of orthogonal polynomials,

$$u_p(x) = \sum_{j=0}^{p} a_j T_j(x), \tag{18}$$

where $T_j(x)$ is the Tchebycheff polynomial of degree j in the variable x with the leading coefficient unity.* However, for theoretical discussions, the second form is much the more convenient because of the simplification brought about by the orthogonal property. The b_{pj} are linear functions of the a_j—in particular, the coefficient of highest degree b_{pp} is identical with a_p. It is the coefficients a_j which are obtained in the forward section of the standard Doolittle method for the solution of the normal equations. The concentration on the orthogonal coefficients must not be taken to imply that the actual fitting is to be done by calculating the orthogonal polynomials at the points of observation—in fact, the usual scheme based on power moments should be employed. The relation between the power series and the orthogonal representation has been fully discussed by Guest (1950) and by Hayes & Vickers (1951).

The variance of a_j is given by the formula

$$\operatorname{var} a_j = \operatorname{var} y / \sum_i T_j^2(x_i). \tag{19}$$

The values x_i are, in the representation used here, replaced by the smoothed set $X_n(\epsilon)$. Using equations (7), (8) and (10), and the standard expressions for $\tau_{ln}(\epsilon) = n^{-l+1} T_{ln}(\epsilon)$, it can be shown that $\sum_i T^2(x_i)$ is given by an approximate equation of the form

$$\sum_i T_j^2(x_i) \doteq \phi_n^{-2j} n^{2j+1} f_j (1 - n^{-2} g_j), \tag{20}$$

where higher powers of n^{-2} have been neglected. Similarly, for the grouped set,

$$\sum_i T_{jN}^2(x_{Ni}) \doteq (r^2 \phi_n^{-1})^{2j} N^{2j+1} f_j (1 - N^{-2} g_j). \tag{21}$$

* It should be noted that $T_j(x)$ and $T_j(\epsilon)$ are different sets of polynomials. The use of the symbols T_j to denote any set of orthogonal polynomials is convenient, but it is realized that it may be confusing in certain cases. The two sets of orthogonal polynomials are here distinguished by the use of different symbols x and ϵ for the variables.

In these formulae f_j and g_j are functions of κ_2 and κ_3, but not of n or N. It will be assumed in the subsequent discussion that the number of observations is so large that the term in n^{-2} may be omitted.

Since y_{Ni} is the sum of r observations y_i,

$$\operatorname{var} y_{Ni} = r \operatorname{var} y_i. \tag{22}$$

If the curve fitted to the N grouped observations is written as

$$u_{pN}(x) = \sum_j a_{jN} T_{jN}(x), \tag{23}$$

then $$\operatorname{var} a_{jN} = \operatorname{var} y_{Ni} / \sum_i T_{jN}^2(x_{Ni})$$

and $$(\operatorname{var} a_j)/(\operatorname{var} a_{jN}) \doteq r^{2(j-1)}\{1 - N^{-2} g_j\}.$$

The efficiency $\eta(a_{jN})$ of the grouped estimate will then be given by

$$\eta(a_{jN}) \doteq 1 - N^{-2} g_j. \tag{24}$$

In practice the g_j are too complicated to be represented by an explicit formula, but it is fairly straightforward to calculate them for given values of κ_2 and κ_3. The expressions for $\eta(a_{jN})$ are listed in Table 1 for polynomials of the first, second and third degrees, and for $\kappa_2^2 = 0(0{\cdot}5)\,1{\cdot}0$, $\kappa_3 = -1{\cdot}0(0{\cdot}5)\,1{\cdot}0$.

From these expressions the efficiencies for any value of N can be determined. The larger the value of N the higher the efficiencies, but also the longer the time required for the calculations. Some compromise must be made, and it appears that suitable values of N lie in the range 9–12 for a second degree curve, and in the range 16–21 for a third-degree curve. For a first-degree curve five groups are sufficient. The efficiencies for these suggested values of N are shown in Table 2.

Bias of the grouped estimates

If the polynomial is of the second or third degree, grouping will usually lead to bias in the estimates. The origin of the bias is seen most clearly by considering the power-series representation. Denoting 'true' values by a prime superscript, the expectation of the observation corresponding to a given value x_i is

$$E(y_i) = \sum_j b'_{pj} x_i^j.$$

Hence $$E(y_{Ni}) = \sum_r E(y_i) = \sum_j b'_{pj} \sum_r x_i^j,$$

the suffix r attached to the summation sign indicating that the sum is taken over the particular group. Now

$$\sum_r x_i^j \doteq r^{-j+1} (\sum_r x_i)^j,$$

unless j is 0 or 1, and so

$$E(y_{Ni}) \doteq \sum_j (r^{-j+1} b'_{pj})\, x_{Ni}^j. \tag{25}$$

The true value of the coefficient for the grouped curve is

$$b'_{pjN} = r^{-j+1} b'_{pj}.$$

Thus (25) becomes

$$E(y_{Ni}) \doteq \sum_j b'_{pjN} x_{Ni}^j. \tag{26}$$

Table 1. *Efficiencies $\eta(a_{jN})$ of the coefficients a_{jN}*

κ_2^2			0			0·5			1·0		
κ_3	−1·0	−0·5	0	0·5	1·0	−1·0	−0·5	0	0·5	1·0	−1·0	−0·5	0	0·5	1·0
g_1	0·82	0·86	1·00	1·29	1·83	0·92	0·98	1·13	1·44	1·98	1·01	1·08	1·25	1·57	2·12
g_2	1·81	2·87	5·00	8·33	12·9	2·88	4·08	6·26	9·61	14·3	3·83	5·20	7·50	11·0	15·9
g_3	2·81	6·49	14·0	25·5	41·6	6·59	10·8	18·4	30·1	46·3	9·68	14·7	22·8	34·6	50·9

$\eta(a_{jN}) \doteqdot 1 - N^{-2}g_j$.

Table 2. *Percentage efficiencies for suggested values of N*

κ_2^2			0			0·5			1·0		
κ_3	−1·0	−0·5	0	0·5	1·0	−1·0	−0·5	0	0·5	1·0	−1·0	−0·5	0	0·5	1·0
						1st degree polynomial									
$\eta(a_{1N})$ $N=5$	96·7	96·6	96·0	94·8	92·7	96·3	96·1	95·5	94·2	92·1	96·0	95·7	95·0	93·7	91·5
						2nd degree polynomial									
$\eta(a_{2N})$ $N=9$	97·8	96·5	93·8	89·7	84·1	96·4	95·0	92·3	88·1	82·3	95·3	93·6	90·7	86·5	80·3
$N=12$	98·7	98·0	96·5	94·2	91·0	98·0	97·2	95·7	93·3	90·1	97·3	96·4	94·8	92·4	88·9
						3rd degree polynomial									
$\eta(a_{3N})$ $N=16$	98·9	97·5	94·5	90·0	83·7	97·4	95·8	92·8	88·3	81·9	96·2	94·3	91·1	86·5	80·1
$N=21$	99·4	98·5	96·8	94·2	90·6	98·5	97·6	95·8	93·2	89·5	97·8	96·7	94·8	92·1	88·5

Table 3. *Quantities B_{jk} for the calculation of the bias in a_{kN}*

κ_2^2			0			0·5				1·0				
κ_3	−1·0	−0·5	0	0·5	1·0	−1·0	−0·5	0	0·5	1·0	−1·0	−0·5	0	0·5	1·0
B_{21}	0	0	0	0	0	0·035	0·051	0·069	0·091	0·117	0·052	0·074	0·099	0·130	0·166
B_{31}	0·055	0·057	0·072	0·101	0·143	0·075	0·079	0·096	0·126	0·172	0·096	0·102	0·120	0·151	0·197
B_{22}	−0·065	−0·035	0	0·040	0·086	−0·063	−0·033	0	0·035	0·074	−0·061	−0·031	0	0·032	0·064
B_{32}	0	0	0	0	0	0·021	0·036	0·053	0·069	0·083	0·033	0·054	0·076	0·099	0·118
B_{23}	0	0	0	0	0	0·001	0·002	0	−0·005	−0·015	0·002	0·002	0	−0·006	−0·017
B_{33}	−0·051	−0·027	0	0·029	0·059	−0·050	−0·025	0	0·027	0·056	−0·050	−0·025	0	0·025	0·052

Because of this inequality the estimates b_{pjN} obtained from the ordinary normal equations

$$\sum_i (y_{Ni} - \sum_j b_{pjN} x_{Ni}^j)\, x_{Ni}^k = 0 \tag{27}$$

will in fact be biased estimates.

Again, it is more convenient (although the arithmetic is still very complicated) to calculate the bias in the orthogonal coefficients rather than in the power-series coefficients. The bias in a_{kN} is defined as the difference of the expectation from the true value,

$$E(a_{kN}) - a'_{kN}.$$

The bias turns out to be, to order N^{-2}, of the form

$$N^{-2} \sum_{j=0}^{p} g_{jk} (\phi_n^{-1} N r^2)^{j-k} a'_{jN}, \tag{28}$$

where the g_{jk} are functions of κ_2 and κ_3. The true values a'_{jN} are of course unknown, but an estimate of the bias can be obtained by substituting the calculated values a_{jN} for a'_{jN} in (28).

From (11) and the standard expressions for $\tau_{jN}(\epsilon)$, the range of $\xi_N(\epsilon)$ is $(N-1)(\kappa_{1N} + \frac{1}{6}\kappa_{3N})$. So, from (12), $\phi_n^{-1} N r^2$ is approximately the range of $X_N(\epsilon)$ and hence also of x_{Ni}—the difference between the greatest and least values. Thus the estimate of bias can be written as

$$N^{-2} \sum_{j=0}^{p} g_{jk} a_{jN} (Ra\, x_{Ni})^{j-k}. \tag{29}$$

The ratio of bias to standard error will be more useful than the actual bias in deciding whether the grouping will be satisfactory. The standard error of a_{kN} is, from (21),

$$(\phi^{-1} N r^2)^{-k} N^{-\frac{1}{2}} f_k^{-\frac{1}{2}} (\text{S.E.}\, y_N),$$

and so the ratio can be put in the form

$$\frac{\text{Bias}\, a_{kN}}{\text{S.E.}\, a_{kN}} \doteq \frac{1}{\sigma_N} N^{-\frac{1}{2}} \sum_j B_{jk} a_{jN} (Ra\, x_{Ni})^j, \tag{30}$$

where σ_N is the standard error of a grouped observation and the B_{jk} are functions of κ_2 and κ_3. B_{0k} and B_{1k} vanish, and so for the third-degree polynomial the bias depends only on a_2 and a_3, while for the second-degree polynomial the bias depends only on a_2. As was the case with the g_j, the explicit formulae for the B_{jk} are excessively complicated, but it is possible to obtain numerical values for selected κ_2 and κ_3. Some of these values are listed in Table 3.

It will often be necessary to ascertain before proceeding with the calculations whether the grouping will give rise to significant bias. In terms of the observations before grouping equation (30) becomes

$$\frac{\text{Bias}\, a_{kN}}{\text{S.E.}\, a_{kN}} \doteq \frac{n^{\frac{1}{2}}}{\sigma} N^{-2} \sum_j B_{jk} A_{jn}, \tag{31}$$

where σ is the standard error of an observation and A_{jn} is written for $a_{jn}(Ra\, x_i)^j$. The values A_{jn} can be estimated in the following way. If the five values of y_i spaced at intervals of one-quarter of the range of x_i are denoted by $y(+\frac{1}{2})$, $y(+\frac{1}{4})$, $y(0)$, $y(-\frac{1}{4})$, $y(-\frac{1}{2})$, then

$$\left.\begin{aligned} A_{2n} &\doteq 2[y(+\tfrac{1}{2}) + y(-\tfrac{1}{2}) - 2y(0)], \\ A_{3n} &\doteq \tfrac{16}{3}[\{y(+\tfrac{1}{2}) - y(-\tfrac{1}{2})\} - 2\{y(+\tfrac{1}{4}) - y(-\tfrac{1}{4})\}]. \end{aligned}\right\} \tag{32}$$

These are actually the values for a curve passing through the five points. The estimate of A_{2n} will usually be reasonably accurate, the estimate of A_{3n} much less so. If the scatter of the observations is large an average value of y in the particular region should be taken. σ can be estimated roughly from the scatter of the observations. κ_2 and κ_3 can be estimated from the values of x corresponding to values of ϵ of $\pm\frac{1}{2}(n-1)$, $\pm\frac{1}{4}(n-1)$, 0, by the formulae

$$\left.\begin{aligned}\kappa_2 &\doteqdot 2\frac{x(+\tfrac{1}{2})+x(-\tfrac{1}{2})-2x(0)}{x(+\tfrac{1}{2})-x(-\tfrac{1}{2})},\\ \kappa_3 &\doteqdot \frac{8}{3}\frac{x(+\tfrac{1}{2})-x(-\tfrac{1}{2})-2\{x(+\tfrac{1}{4})-x(-\tfrac{1}{4})\}}{x(+\tfrac{1}{2})-x(-\tfrac{1}{2})}.\end{aligned}\right\} \tag{33}$$

ϵ is the number specifying the position in the sequence.

ILLUSTRATIVE EXAMPLE

In Table 4 is listed a set of 67 observations of the mechanical equivalent of heat as a function of temperature (Jaeger & von Steinwehr, 1921). Suitable constants have been subtracted from the original observations to reduce their magnitudes. A rough check on the order of the biases for $N = 17$ will first be made.

The range of x_i is from $+30$ to -15. Hence values of y in the neighbourhood of $x = +30$, $+19$, $+7$, -4, -15 are required for substitution in formulae (32). Average values in these regions are of the order of

$$y(+\tfrac{1}{2}) = 110, \quad y(+\tfrac{1}{4}) = 80, \quad y(0) = 50, \quad y(-\tfrac{1}{4}) = 120, \quad y(-\tfrac{1}{2}) = 290,$$

and so

$$A_{2n} \doteqdot +600, \quad A_{3n} \doteqdot -550.$$

These values are of course quite rough, but will suffice for the estimation of the biases.

The standard error σ, estimated from the scatter of the observations, is about 30. Since there are 67 observations, the values of x_i for the observations numbered 1, $17\frac{1}{2}$, 34, $50\frac{1}{2}$, 67 are required for the estimation of κ_2 and κ_3 from formulae (33). These values are:

$$x(+\tfrac{1}{2}) = 29{\cdot}6, \quad x(+\tfrac{1}{4}) = 11{\cdot}0, \quad x(0) = 1{\cdot}1, \quad x(-\tfrac{1}{4}) = -6{\cdot}5, \quad x(-\tfrac{1}{2}) = -15{\cdot}2,$$

and so

$$\kappa_2 \doteqdot 0{\cdot}54, \quad \kappa_2^2 \doteqdot 0{\cdot}3, \quad \kappa_3 \doteqdot 0{\cdot}6.$$

Hence, from Table 3,

$$\begin{aligned} B_{21} &\sim 0{\cdot}07, \quad B_{22} \sim 0{\cdot}04, \quad B_{23} \sim 0{\cdot}00,\\ B_{31} &\sim 0{\cdot}12, \quad B_{32} \sim 0{\cdot}06, \quad B_{33} \sim 0{\cdot}03.\end{aligned}$$

From equation (31)

$$\frac{\text{Bias}\, a_{1N}}{\text{S.E.}\, a_{1N}} = \frac{\sqrt{67}}{30}\frac{1}{289}(0{\cdot}07 \times 600 - 0{\cdot}12 \times 550) = -0{\cdot}02.$$

The ratios for the other coefficients will be of the same order, and so the bias with $N = 17$ will be negligible. It will be observed that this extremely small bias in the estimates occurs because A_{2n} and A_{3n} are of opposite sign, but even if they were of the same sign the ratio would still only be 0·10.

The grouped values are shown in Table 4 (*b*). The central group contains only three observations, and so the sums for this group are increased by 4/3. The calculations using the Doolittle scheme have been carried out both for the original set of observations and for the

grouped set. Table 4 (*c*) gives the calculated values of $\Sigma T_j^2(x)$ and the efficiencies of the grouped estimates. These are compared with the efficiencies as predicted by Table 1 for $\kappa_2^2 = 0{\cdot}3$, $\kappa_3 = 0{\cdot}6$. The values given by Table 1 are quite close to the true values—in fact, closer than might have been expected considering the inaccuracy of the estimate of κ_3. Table 4 (*d*) gives the orthogonal coefficients, and the power-series coefficients for the third-degree curves. The biases are seen to be completely negligible, as was predicted. There will, of course, be random differences between the two sets of coefficients, and these completely swamp any bias effects in this case.

Table 4. *Illustrative example*

(*a*) *Original observations* ($n = 67$)

x	y	x	y	x	y	x	y
+29·60	108	+11·49	88	−0·05	119	− 7·13	101
+28·36	102	+10·60	57	−0·25	106	− 7·40	165
+26·96	103	+ 9·19	58	−1·17	104	− 7·41	169
+25·79	74	+ 9·15	79	−1·18	117	− 7·94	187
+25·56	25	+ 7·76	66	−1·45	104	− 8·39	153
+24·34	103	+ 6·33	41	−1·58	97	− 8·55	175
+23·09	104	+ 6·24	71	−1·58	104	− 8·67	202
+19·41	66	+ 5·55	50	−2·59	133	− 9·98	176
+17·19	35	+ 5·36	70	−2·67	90	−11·05	207
+16·64	173	+ 4·82	60	−3·61	137	−11·19	200
+15·79	85	+ 4·11	66	−3·69	130	−11·49	230
+15·75	83	+ 3·96	76	−5·01	118	−12·59	254
+14·39	69	+ 3·24	71	−5·12	114	−13·62	259
+14·32	113	+ 2·53	80	−6·00	201	−13·85	245
+13·98	45	+ 1·80	83	−6·03	114	−14·35	315
+12·54	68	+ 1·41	80	−6·97	137	−15·25	291
		+ 1·15	89				
		+ 1·13	96				
		+ 0·01	99				

(*b*) *Grouped observations* ($N = 17$)

x	y	x	y	x	y	x	y
+110·71	387	+40·43	282	− 2·65	446	−29·88	622
+92·40	298	+25·88	228	− 7·20	438	−35·59	706
+65·37	376	+18·25	272	−14·98	475	−46·32	891
+55·23	295	+ 8·98	314	−24·12	566	−57·07	1110
		+ 3·05	379				

Table 4 (*cont.*)

(*c*) *Efficiencies* $\eta(a_{jN})$

j	ΣT_{jn}^2	ΣT_{jN}^2	$\eta = \Sigma T_{jN}^2 / r^{2j-1} \Sigma T_{jn}^2$	η (Table 1)
1	9265	36861	0·9946	$1 - 1{\cdot}46/289 = 0{\cdot}9950$
2	$143{\cdot}66 \times 10^4$	8922×10^4	0·970	$1 - 9{\cdot}9/289 = 0{\cdot}966$
3	$1{\cdot}913 \times 10^8$	1780×10^8	0·909	$1 - 31{\cdot}1/289 = 0{\cdot}892$

(*d*) *Orthogonal coefficients a and power-series coefficients b*

j	a_{jn}	a_{jN}	$r^{j-1}a_{jN}$	b_{3jn}	b_{3jN}	$r^{j-1}b_{3jN}$
0	—	—	—	$92{\cdot}4 \pm 4{\cdot}4$	372·18	93·0
1	$-3{\cdot}59 \pm 0{\cdot}25$	−3·59	−3·59	$-5{\cdot}57 \pm 0{\cdot}48$	−5·60	−5·60
2	$+0{\cdot}261 \pm 0{\cdot}020$	+0·0661	+0·264	$+0{\cdot}407 \pm 0{\cdot}040$	+0·1002	+0·401
3	$-0{\cdot}0074 \pm 0{\cdot}0018$	−0·000451	−0·0072	$b_{33} = a_3$		

The standard error of an observation may be estimated from the usual formulae

$$\{\Sigma v_i^2/(n-p-1)\}^{\frac{1}{2}} \quad \text{and} \quad \{\Sigma v_{iN}^2/(N-p-1)\}^{\frac{1}{2}}$$

for the ungrouped and grouped cases respectively. In both cases the sum of the squares of the residuals can be found from the formula

$$\Sigma v^2 = \Sigma y^2 - \Sigma a_j^2 \Sigma T_j^2.$$

In the present example the estimate of σ in the ungrouped case is 24, and the estimate of σ_N in the grouped case is 37. These should be in the ratio 1:2. It is not clear whether the departure from this ratio is significant.

Weighting of grouped observations

Often it is not possible to find a suitable pair of values N, r, such that $n = Nr$. More usually,

$$n = Nr + \nu, \tag{34}$$

where ν is small. The ν additional observations should be included in ν groups near the centre of the range of x_i, and in forming these groups the sums of the x_i and y_i values must be multiplied by $r/(r+1)$ to bring them to the same scale as the other groups. If ν is negative the factor is $r/(r-1)$.

These ν groups should strictly be given a weight $(r \pm 1)/r$ in forming the moments and the sums of the powers. This, however, greatly increases the time required to evaluate these sums, and the calculations can be done much more rapidly if the weights are omitted. The omission of the weights does not lead to any increase in the bias, but it does produce a drop in the efficiency. A detailed investigation, too long to be reproduced here, indicates a drop in efficiency by a factor of the order of $1 - \nu r^{-2} N^{-1}$, which will almost always be negligible. Hence it is recommended that the grouped observations be left unweighted.

If the original set of observations y_i have different weights w_i, then Σw_i replaces n in formula (34), and the observation of weight w_i is just regarded as equivalent to w_i observations of unit weight. The observations are divided into N groups each having the same Σw_i.

Step-function methods

For the first-degree polynomial the use of step functions provides the most rapid method of determining the line which fits the observations. Since in practice the majority of the curves to be fitted will be of the first degree, a full discussion of the use of step functions in these cases will now be presented.

If the scale factor is removed, the independent variable may be represented as

$$\xi(\epsilon) = \kappa_1\epsilon + \kappa_2 n^{-1}(\epsilon^2 - \tfrac{1}{12}n^2) + 2\kappa_3 n^{-2}(\epsilon^3 - \tfrac{3}{20}n^2\epsilon), \tag{35}$$

terms of the order of n^{-2} being neglected. Since κ_1 is $1 - \frac{1}{5}\kappa_3$, this can be regrouped as

$$\xi(\epsilon) = \epsilon(1 - \tfrac{1}{2}\kappa_3) + 2\kappa_3 n^{-2}\epsilon^3 + \kappa_2 n^{-1}(\epsilon^2 - \tfrac{1}{12}n^2). \tag{36}$$

If $w_1(\epsilon)$ is any function of ϵ for which $\Sigma w_1(\epsilon) = 0$, then

$$b_1 = \Sigma w_1(\epsilon)\, y(\epsilon) / \Sigma w_1(\epsilon)\, \xi(\epsilon)$$

will provide an estimate of the slope of the straight line which fits the observations. The standard error of this estimate is given by

$$\sigma^2(y)/\sigma^2(b_1) = \{\Sigma w_1(\epsilon)\, \xi(\epsilon)\}^2 / [\Sigma w_1^2(\epsilon)]. \tag{37}$$

For a step function, $\Sigma w_1(\epsilon) f(\epsilon)$ is of the form

$$\left[q_1\left(\sum_{0,\frac{1}{2}}^{\frac{1}{2}(n-1)} - \sum_{0,\frac{1}{2}}^{\frac{1}{2}(a_1-1)}\right) + q_2\left(\sum_{0,\frac{1}{2}}^{\frac{1}{2}(a_1-1)} - \sum_{0,\frac{1}{2}}^{\frac{1}{2}(a_2-1)}\right) + \ldots + q_m\left(\sum_{0,\frac{1}{2}}^{\frac{1}{2}(a_{m-1}-1)} - \sum_{0,\frac{1}{2}}^{\frac{1}{2}(a_m-1)}\right)\right][f(\epsilon) - f(-\epsilon)], \tag{38}$$

the numbers $\frac{1}{2}(a_j - 1)$ being the values of ϵ at the ends of the steps. Neglecting terms of the order of n^{-2},

$$\left.\begin{aligned} \sum_{0,\frac{1}{2}}^{\frac{1}{2}(a-1)} \{\epsilon^j - (-\epsilon)^j\} &= \{n^{j+1}/(j+1)\, 2^j\}\, \alpha^{j+1}, \quad j \text{ odd}, \\ &= 0, \quad j \text{ even}, \end{aligned}\right\} \tag{39}$$

where $n\alpha = a$. Hence, from (38),

$$\Sigma w_1(\epsilon)\, \epsilon^j = \{n^{j+1}/(j+1)\, 2^j\} \sum_{k=1}^{m} q_{m-k+1}(\alpha_{m-k}^{j+1} - \alpha_{m-k+1}^{j+1}) \tag{40}$$

when j is odd, and

$$\Sigma w_1^2(\epsilon) = n \sum_{k=1}^{m} q_{m-k+1}^2(\alpha_{m-k} - \alpha_{m-k+1}). \tag{41}$$

For the equally spaced case it has been shown (Guest, 1954) that the estimate b_1 is of smallest standard error when the steps are of equal size and the weights are $m, m-1, \ldots, 1$. Using then the values

$$q_{m-k+1} = k, \quad \alpha_{m-k} = (2k+1)/(2m+1),$$

in (40),

$$\Sigma w_1(\epsilon)\, \epsilon = \{2n^2/(2m+1)^2\}\, \Sigma k^2, \tag{42}$$

$$\Sigma w_1(\epsilon)\, \epsilon^3 = \{n^4/2(2m+1)^4\}\, \{\Sigma k^2 + 4\Sigma k^4\}, \tag{43}$$

while from (41)

$$\Sigma w_1^2(\epsilon) = \{2n/(2m+1)\}\, \Sigma k^2. \tag{44}$$

Substituting these values in (37) and inserting the standard expressions for Σk^j, it is found that

$$\sigma^2(y)/\sigma^2(b_1) = [\tfrac{1}{12}n^3][1-(2m+1)^{-2}][1-\tfrac{1}{5}\kappa_3\{1+(2m+1)^{-2}\}]^2,$$

while for the least-squares estimate $b_1' \equiv a_1'$, (19) and (20) give

$$\sigma^2(y)/\sigma^2(b_1') = n^3 f_1.$$

The explicit expression for f_1 is

$$12f_1 = 1-\tfrac{2}{5}\kappa_3+\tfrac{2}{35}\kappa_3^2+\tfrac{1}{15}\kappa_2^2.$$

Therefore the efficiency $\eta(b_1)$ of the estimate b_1 is given by

$$\eta(b_1) = [1-N^{-2}][1-\tfrac{1}{5}\kappa_3(1+N^{-2})]^2/12f_1, \tag{45}$$

where $N = 2m+1$ is the number of groups.

The function $\eta(b_1)$ is tabulated in Table 5 for values of N equal to 3, 5, 7, ∞. The efficiencies when $N = 3$ are perhaps a little low, but the value $N = 5$, corresponding to double-step functions, gives satisfactory efficiencies.

Table 5. *Percentage efficiencies in first-degree step-function methods*

κ_2^2	0					0·5					1·0				
κ_3	−1·0	−0·5	0	0·5	1·0	−1·0	−0·5	0	0·5	1·0	−1·0	−0·5	0	0·5	1·0
$N=3$	91·1	90·4	88·9	86·3	81·8	89·1	88·0	86·0	82·9	77·9	87·1	85·7	83·3	79·7	74·3
$N=5$	96·1	96·4	96·0	94·6	91·6	94·0	93·8	92·9	90·9	87·2	91·9	91·3	90·0	87·5	83·2
$N=7$	97·5	98·0	98·0	97·0	94·4	95·3	95·4	94·8	93·2	89·9	93·2	92·9	91·8	89·7	85·7
$N=\infty$	98·8	99·6	100	99·5	97·4	96·6	97·0	96·8	95·6	92·7	94·5	94·5	93·8	91·9	88·4

An account has been given in two earlier papers (Guest, 1952, 1953*b*) of the use of single-step functions in the fitting of higher degree polynomials. Step-function methods are rapid and give unbiased estimates, but the efficiencies are rather low if the departure from uniform spacing is at all pronounced. An investigation into the use of double-step functions has since been carried out, but it appears that the improvement in efficiency is only slight for markedly non-uniform spacing. One weakness of the earlier method is that different functions were used for the second- and third-degree curves. It has since been found that if the steps in the third-degree equation are weighted in the ratio 2:1, the same functions can be used for the second- and third-degree curves without any great drop in efficiency. The recommended step functions are now given by the expressions below:

Zero degree:	Σn
First degree:	$\Sigma n_{11}+\Sigma n_{11}'-\Sigma n;$
Second degree:	$\Sigma n_{21}+\Sigma n_{22}-\Sigma n_{22}'-\Sigma n_{21}'+\Sigma n;$
Third degree:	$2\Sigma n_{31}+\Sigma n_{32}-\Sigma n_{33}-\Sigma n_{33}'+\Sigma n_{32}'+2\Sigma n_{31}'-2\Sigma n.$

The observations are supposed numbered from 1 to n, and Σn_{jk} signifies the sum for all the observations from 1 to n_{jk}. The values n_{jk} for the various step functions are the integers nearest to the numbers shown below:

$$n_{11}\text{: } 0{\cdot}33n; \quad n_{21}\text{: } 0{\cdot}11n; \quad n_{22}\text{: } 0{\cdot}25n;$$
$$n_{31}\text{: } 0{\cdot}07n; \quad n_{32}\text{: } 0{\cdot}15n; \quad n_{33}\text{: } 0{\cdot}44n;$$
$$n_{jk}' = n-n_{jk}.$$

Conclusion

To obtain an idea of the times required for the various methods, the calculations were performed by each method for the example given in Table 4. Table 6 gives a summary of the times required to fit polynomials of different degrees (including the checking of the calculations) by the various methods, together with the efficiencies and the estimates of the coefficients a_p. It will be seen that for the first-degree polynomial the step-function method is by far the most rapid one. For the second-degree polynomial the step function and the least-squares grouped methods require about the same time, while for the third-degree polynomial the least-squares grouped method is the most rapid.

Table 6. *Results for selected example* (67 *observations*)

	1st Degree		2nd Degree			3rd Degree		
	Time (min.)	Effic. $\eta(a_1)$	Time (min.)	Effic. $\eta(a_2)$	$a_2(\times 10^2)$	Time (min.)	Effic. $\eta(a_3)$	$a_3(\times 10^3)$
Least-squares	58	1·000	158	1·000	$26{\cdot}1 \pm 2{\cdot}3$	246	1·000	$-7{\cdot}4 \pm 1{\cdot}8$
Least-squares grouped	29	0·936	61	0·925	25·8	100	0·909	−7·2
Single-step functions	16	0·825	60	0·831	28·4	128	0·725	−7·9
Double-step functions	17	0·898	67	0·835	28·9	145	0·763	−7·1

The general conclusion to be drawn from the work described here is that each method has its disadvantages. The full least-squares method is efficient and without bias, but it takes a very long time, especially if the computor is not familiar with the technique. The least-squares grouped method is rapid and the efficiency is high, but there may be bias in the estimates. The single-step function method is rapid and of fair efficiency, and there is no bias, but the standard errors cannot be estimated in any simple way.

Perhaps the following statements will provide a satisfactory guide in the selection of the appropriate method for the fitting of a second- or third-degree polynomial. If the scatter is large the least-squares grouped method should be used, since the bias is not then likely to be of importance. Cases will also arise in which it is possible to remove the greater part of the variation in the dependent variable by the use of approximate values of the coefficients. In particular, if the approximate contributions $b_2x^2 + b_3x^3$ are subtracted before grouping, the corresponding least-squares coefficients obtained from the grouped values will be so small that the bias due to grouping will be negligible. The efficiency is unaffected by this process. If this procedure is not convenient and the scatter is small the step-function method may be used, but if accurate estimates of standard error are also required there is no alternative but to fit a least-squares curve to the full set of observations.

REFERENCES

Guest, P. G. (1950). *Phil. Mag.* [7], **41**, 124.
Guest, P. G. (1952). *Aust. J. Sci. Res.* A, **5**, 238.
Guest, P. G. (1953*a*). *Aust. J. Phys.* **6**, 131.
Guest, P. G. (1953*b*). *Aust. J. Phys.* **6**, 361.
Guest, P. G. (1954). *Biometrika*, **41**, 62.
Hayes, J. G. & Vickers, T. (1951). *Phil. Mag.* [7], **42**, 1387.
Jaeger, W. & von Steinwehr, H. (1921). *Ann. Phys., Lpz.*, **64**, 305.

ON THE JOINT DISTRIBUTION OF THE CIRCULAR SERIAL CORRELATION COEFFICIENTS

BY G. S. WATSON†

The Australian National University, Canberra, A.C.T.

1. INTRODUCTION

Quenouille (1949) applied Koopmans's method (1942) to the problem of finding the exact joint distribution of the serial correlation coefficients in the null case. Madow's (1945) device was used to derive the non-null joint density function. For odd sample sizes, he gave an explicit formula for the exact density function. Since the direct smoothing of this function is very difficult, Quenouille conjectured a smooth approximation to the exact density and supported his conjecture by various arguments about the expected form of the result. Jenkins (1954), following Dixon's (1944) method, has found a smoothed density for the first two serial correlation coefficients which, unlike Quenouille's density, has the correct moments up to order n. These smoothed densities were derived for the case of no mean corrections.‡

In §2 of the present paper, von Neumann's (1941) method is generalized to give integral equations for the joint density of the serial correlations. From these the relationship of the null and non-null densities is easily derived. In §3 the exact density of the circular serial correlations is obtained for arbitrary sample sizes, and a fuller discussion is given of the interesting summation rule involved. In §4 an attempt is made to derive an approximate null density when mean corrections are made by a new smoothing method. The form found differs only slightly from Quenouille's and is subject to the same weaknesses.

2. SOME GENERAL RESULTS

The basic problem is to find the joint probability density function of

$$r_j = \frac{\mathbf{x}'\mathbf{A}_j\mathbf{x}}{\mathbf{x}'\mathbf{x}} \quad (j = 1, \ldots, q), \tag{1}$$

where the elements of the column vector $\mathbf{x}$ are independent normal variates with zero mean and unit variance. It does not seem possible to make any progress unless the $\mathbf{A}_j$ form a commutative set of matrices. In the present applications this means that serial correlations with a modified definition must be considered. With this assumption the problem may be reduced to that of the joint distribution of

$$r_j = \frac{\lambda_0^{(j)} w_0 + \lambda_1^{(j)} w_1 + \ldots + \lambda_m^{(j)} w_m}{w_0 + w_1 + \ldots + w_m} \quad (j = 1, \ldots, q), \tag{2}$$

† Most of the work reported in this paper was done while the author was a Research Officer in the Department of Applied Economics, University of Cambridge, and formed part of a thesis issued by the Institute of Statistics, University of North Carolina, Mimeograph Series no. 49.

‡ Recently H. E. Daniels (1956) has derived the approximate joint density, with and without mean corrections, for moderate sample sizes, n, and q not comparable with n. [See also the paper by G. M. Jenkins (1956). These two papers are printed on pp. 169 and 186 below. ED.]

where $\lambda_0^{(j)}, \lambda_1^{(j)}, \ldots, \lambda_m^{(j)}$ are the distinct latent roots of $\mathbf{A}_j$ with multiplicities $p_0, p_1, \ldots, p_m$ and $w_0, w_1, \ldots, w_m$ are independent gamma variables of orders $p_0, p_1, \ldots, p_m$. The multiplicities of the roots of $\mathbf{A}_j$ associated with any latent vector must be independent of j to obtain the reduction (2).

Writing $r_j = u_j/v\,(j = 1, \ldots, q)$, Pitman's theorem (1937) gives immediately

$$E(r_j^{k_1} r_j^{k_2} r_j^{k_3} \ldots) = \frac{E(u_j^{k_1} u_j^{k_2} u_j^{k_3} \ldots)}{E(v^{k_1+k_2+k_3+\cdots})}, \tag{3}$$

since v is independent of the set of values $r_1, r_2, \ldots, r_q$. Writing $f(r_1, \ldots, r_q)$ for the joint density function of $r_1, \ldots, r_q$ and $g(v) = v^{P-1}e^{-v}/\Gamma(P)$ with $P = \sum_{i=0}^{m} p_i$, the joint moment generating function of $u_1, u_2, \ldots, u_q$ is easily seen to be given by

$$\begin{aligned} E\{\exp(t_1u_1 + t_2u_2 + \ldots + t_qu_q)\} &= E\{\exp[v(r_1t_1 + r_2t_2 + \ldots + r_qt_q)]\} \\ &= \int_{D_q}\ldots\int dr_1 \ldots dr_q \int_0^\infty dv\, e^{v\Sigma r_i t_i} g(v) f(r_1, \ldots, r_q) \\ &= \int_{D_q}\ldots\int \frac{f(r_1, \ldots, r_q)}{\left(1 - \sum_{i=1}^{q} r_i t_i\right)^P} dr_1 \ldots dr_q, \end{aligned} \tag{4}$$

where D_q is the domain of joint variation of $r_1, r_2, \ldots, r_q$, defined below. But the left-hand side of (4) may be evaluated directly to give the integral equation for $f(r_1, \ldots, r_q)$

$$\prod_{i=0}^{m} (1 - \lambda_i^{(1)}t_1 - \ldots - \lambda_i^{(q)}t_q)^{-p_i} = \int_{D_q}\ldots\int \frac{f(r_1, \ldots, r_q)}{(1 - \Sigma r_i t_i)^P} dr_1 \ldots dr_q. \tag{5}$$

This is a generalization of von Neumann's (1941) integral equation for the density function of a single ratio. Using the results of von Neumann & Morgenstern (1947), the domain D_q is seen to be the least convex polyhedron enclosing the points $(\lambda_i^{(1)}, \ldots, \lambda_i^{(q)})\,(i = 0, 1, \ldots, m)$ in the q-space of $r_1, r_2, \ldots, r_q$.

The integral equation (5) provides a simple method of deriving the joint density of $r_1, r_2, \ldots, r_q$ when the $\mathbf{x}$ vector has a multivariate normal density function proportional to

$$\exp\{-\tfrac{1}{2}\mathbf{x}'(\mathbf{I} - \delta_1\mathbf{A}_1 - \ldots - \delta_q\mathbf{A}_q)\mathbf{x}\}, \tag{6}$$

where $\delta_1, \ldots, \delta_q$ are such that the density is non-singular. Since

$$\sum_{j=1}^{q} t_j \mathbf{x}'\mathbf{A}_j\mathbf{x} = \frac{\Sigma t_j \mathbf{x}'\mathbf{A}_j\mathbf{x}}{\mathbf{x}'\mathbf{x} - \Sigma\delta_j\mathbf{x}'\mathbf{A}_j\mathbf{x}}(\mathbf{x}'\mathbf{x} - \Sigma\delta_j\mathbf{x}'\mathbf{A}_j\mathbf{x}) = \frac{\Sigma t_j r_j}{1 - \Sigma\delta_j r_j}\mathbf{x}'\left(\mathbf{I} - \sum_1^q \delta_j\mathbf{A}_j\right)\mathbf{x}, \tag{7}$$

where the first factor of (7) is a homogeneous function of degree zero in gamma variables and where the second factor of (7) is the sum of these variates—a gamma-P variate. The argument leading to (5) now gives

$$\prod_{i=0}^{m}\left(1 - \frac{\lambda_i^{(1)}t_1 + \ldots + \lambda_i^{(q)}t_q}{1 - \delta_1\lambda_i^{(1)} - \ldots - \delta_q\lambda_i^{(q)}}\right)^{-p_i} = \int_{D_q}\ldots\int \frac{f_\delta(r_1, \ldots, r_q)\, dr_1 \ldots dr_q}{\left(1 - \frac{\Sigma r_j t_j}{1 - \Sigma r_j\delta_j}\right)^P}, \tag{8}$$

an integral equation for the non-null joint density of $r_1, \ldots, r_q$. By substitution in (8), it is easily seen that

$$f_\delta(r_1, \ldots, r_q) = \prod_{i=0}^{m} (1 - \sum_j \delta_j \lambda_i^{(j)})^{p_i} (1 - \Sigma \delta_j r_j)^{-P} f(r_1, \ldots, r_q), \tag{9}$$

where $f(r_1, \ldots, r_q)$ is the solution of (5), the null density. A similar, but incorrect, extension of Madow's (1945) result to the joint distribution has been noted by Quenouille (1949). (9) shows that the distributions of the multiple serial correlation $r_{q+1(12\ldots q)}$ and the partial serial correlation $r_{1,q+1.2\ldots q}$ are independent of $\delta_1, \ldots, \delta_q$. This fact for $q = 1$ has been noted by Dixon (1944) for the means and variances of these statistics and by Jenkins (1954) for the distributions.

The integral equations may also be used to demonstrate the ultimate normality of the joint distribution when the sample size tends to infinity. By a different method, Hsu (1946) has examined this question for individual ratios.

3. THE EXACT DENSITY FUNCTION

From §2, it is sufficient to consider the null case. Then

$$r_j = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(x_{i+j} - \bar{x})}{\sum_{i=1}^{N} (x_i - \bar{x})^2}, \tag{10}$$

where $x_{N+i} = x_i$ and where $x_1, x_2, \ldots, x_N$ are independent standard normal variables. The canonical form of (10) is

$$r_j = \frac{\sum_{k=1}^{N-1} \cos \frac{2\pi jk}{N} z_k^2}{\sum_{k=1}^{N-1} z_k^2}. \tag{10'}$$

Anderson (1942) has given the exact marginal distributions of the r_j. For our purposes, we will require only to exhibit the distribution of the latent roots $\cos \frac{2\pi jk}{N}$. For any fixed j, $\cos \frac{2\pi j(N-k)}{N} = \cos \frac{2\pi jk}{N}$, so that the roots are equal in pairs with the possible exception of the root, minus one, the least root. Thus the following formulation will be seen to include our problem. Suppose that

$$\left.\begin{aligned} r_1 &= \frac{\lambda w + \lambda_1 w_1 + \ldots + \lambda_n w_n}{w + w_1 + \ldots + w_n}, \\ r_2 &= \frac{\mu w + \mu_1 w_1 + \ldots + \mu_n w_n}{w + w_1 + \ldots + w_n}, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\ r_q &= \frac{\tau w + \tau_1 w_1 + \ldots + \tau_n w_n}{w + w_1 + \ldots + w_n}, \end{aligned}\right\} \tag{11}$$

where $w_1, \ldots, w_n$ and w are independent gamma variables, the first n variables being gamma-1 and the last gamma-P,† where $\lambda_1, \ldots, \lambda_n$ are all greater than λ, and $\mu_1, \ldots, \mu_n$ all greater than

† If N is even, $p = \frac{1}{2}$; if N is odd, $p = 0$; evidently $p + n = \frac{1}{2}N$.

μ and $\tau_1, \ldots, \tau_n$ all greater than τ. (The problem imposes more restrictions, to be stated later, on the coefficients of the w's.) It is then sufficient to find the joint distribution of

$$s_1 = r_1 - \lambda = \frac{\sum_1^n \lambda_i' w_i}{w + \Sigma w_i} = \frac{l}{v}, \quad \ldots, \quad s_q = r_q - \tau = \frac{\Sigma \tau_i' w_i}{w + \Sigma w_i} = \frac{t}{v}. \tag{12}$$

The region of joint variation of $s_1, s_2, \ldots, s_q$ is the least convex polyhedron enclosing the points $(0, 0, \ldots, 0), (\lambda_1', \mu_1', \ldots, \tau_1'), \ldots, (\lambda_n', \mu_n', \ldots, \tau_n')$, a polyhedron in the positive quadrant of the q-space of $s_1, s_2, \ldots, s_q$. It is convenient to drop the primes for the moment.

The joint characteristic function of $l, m, \ldots, t$ and v is

$$\phi(\theta_1, \theta_2, \ldots, \theta_q, \theta_{q+1}) = \prod_{j=1}^{n} (1 - i\theta_1\lambda_j - i\theta_2\mu_j - \ldots - i\theta_q\tau_j - i\theta_{q+1})^{-1} (1 - i\theta_{q+1})^{-p}. \tag{13}$$

The right-hand side of (13) admits a partial fraction expansion (Watson, 1951) so that we have, for $q < n$,

$$\phi(\theta_1, \ldots, \theta_{q+1}) = \sum_{j_1 < j_2 < \ldots < j_q} \sum \ldots \sum \frac{C_{j_1 j_2 \ldots j_q}}{(1 - i\theta_{q+1})^{n-q+p} \prod_{k=1}^{q} (1 - i\lambda_{j_k}\theta_1 - i\mu_{j_k}\theta_2 - \ldots - i\theta_q\tau_{j_k} - i\theta_{q+1})}, \tag{14}$$

where

$$C_{j_1 j_2 \ldots j_q} = \frac{\begin{vmatrix} \lambda_{j_1} & \mu_{j_1} & \ldots & \tau_{j_1} \\ \lambda_{j_2} & \mu_{j_2} & \ldots & \tau_{j_2} \\ \vdots & \vdots & & \vdots \\ \lambda_{j_q} & \mu_{j_q} & \ldots & \tau_{j_q} \end{vmatrix}^{n-q}}{\prod_{j \neq j_1, j_2, \ldots, j_q} \begin{vmatrix} 1 & \lambda_j & \mu_j & \ldots & \tau_j \\ 1 & \lambda_{j_1} & \mu_{j_1} & \ldots & \tau_{j_1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \lambda_{j_q} & \mu_{j_q} & \ldots & \tau_{j_q} \end{vmatrix}}, \tag{15}$$

provided the λ's, μ's, ..., τ's are such that all the C's are finite. The general term on the right-hand side of (14) may be recognized as the joint characteristic function of

$$\left.\begin{aligned} l' &= \lambda_{j_1} w_{j_1} + \lambda_{j_2} w_{j_2} + \ldots + \lambda_{j_q} w_{j_q}, \\ m' &= \mu_{j_1} w_{j_1} + \mu_{j_2} w_{j_2} + \ldots + \mu_{j_q} w_{j_q}, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ t' &= \tau_{j_1} w_{j_1} + \tau_{j_2} w_{j_2} + \ldots + \tau_{j_q} w_{j_q}, \\ v &= w' + w_1 + \ldots + w_n, \end{aligned}\right\} \tag{16}$$

where w' is gamma $n + p - q$ and is independent of $w_{j_1}, w_{j_2}, \ldots, w_{j_q}$. The joint density of these variables may be found *ab initio* to be (writing sign of $x = \operatorname{sgn} x$ so that $|x| = x \operatorname{sgn} x$)

$$\frac{\operatorname{sgn} \begin{vmatrix} \lambda_{j_1} & \ldots & \tau_{j_1} \\ \vdots & & \vdots \\ \lambda_{j_q} & \ldots & \tau_{j_q} \end{vmatrix} e^{-v} v^{n+p-q-1}}{\begin{vmatrix} \lambda_{j_1} & \ldots & \tau_{j_1} \\ \vdots & & \vdots \\ \lambda_{j_q} & \ldots & \tau_{j_q} \end{vmatrix}^{n-q+p} \Gamma(n+p-q)} \begin{vmatrix} 1 & l'/v & m'/v & \ldots & t'/v \\ 1 & \lambda_{j_1} & \mu_{j_1} & \ldots & \tau_{j_1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \lambda_{j_q} & \mu_{j_q} & \ldots & \tau_{j_q} \end{vmatrix}^{n+p-q-1}. \tag{17}$$

Thus, applying the Fourier inversion operator to (14), using (17) and integrating v out from 0 to ∞, the joint density of $s_1, s_2, \ldots, s_q$ is found to be

$$\frac{\Gamma(n+p)}{\Gamma(n+p-q)} \sum_{(j_1,\ldots,j_q)} \cdots \sum_{\epsilon S_q(s_1,\ldots,s_q)} \frac{\begin{vmatrix} 1 & s_1 & s_2 & \cdots & s_q \\ 1 & \lambda_{j_1} & \mu_{j_1} & \cdots & \tau_{j_1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \lambda_{j_q} & \mu_{j_q} & \cdots & \tau_{j_q} \end{vmatrix}^{n+p-q-1} \operatorname{sgn} \begin{vmatrix} \lambda_{j_1} & \cdots & \tau_{j_1} \\ \vdots & & \vdots \\ \lambda_{j_q} & \cdots & \tau_{j_q} \end{vmatrix}}{\begin{vmatrix} \lambda_{j_1} & \cdots & \tau_{j_1} \\ \vdots & & \vdots \\ \lambda_{j_q} & \cdots & \tau_{j_q} \end{vmatrix}^{p} \prod_{j \neq j_1,\ldots,j_q} \begin{vmatrix} 1 & \lambda_j & \cdots & \tau_j \\ 1 & \lambda_{j_1} & \cdots & \tau_{j_1} \\ \vdots & \vdots & & \vdots \\ 1 & \lambda_{j_q} & \cdots & \tau_{j_q} \end{vmatrix}}, \tag{18}$$

since the Jacobian of $l, m, \ldots, t, v$ to $s_1, s_2, \ldots, s_q, v$ is v^q. The terms in the sum (18) depend on the region of the density (17), i.e. the term $(j_1, \ldots, j_q)$ belongs to $S_q(s_1, \ldots, s_q)$, provided $s_1, s_2, \ldots, s_q$ are possible values of $l'/v, \ldots, t'/v$ as defined in (16). This means that $s_1, \ldots, s_q$ must fall in the least convex polyhedron with vertices $(0, 0, \ldots, 0)$, $(\lambda_{j_1}, \mu_{j_1}, \ldots, \tau_{j_1}), \ldots, (\lambda_{j_q}, \mu_{j_q}, \ldots, \tau_{j_q})$. In this region

$$\frac{\begin{vmatrix} 1 & s_1 & s_2 & \cdots & s_q \\ 1 & \lambda_{j_1} & \mu_{j_1} & \cdots & \tau_{j_1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \lambda_{j_q} & \mu_{j_q} & \cdots & \tau_{j_q} \end{vmatrix}}{\begin{vmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & \lambda_{j_1} & \mu_{j_1} & \cdots & \tau_{j_1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \lambda_{j_q} & \mu_{j_q} & \cdots & \tau_{j_q} \end{vmatrix}} \geqslant 0. \tag{19}$$

As Quenouille has pointed out the general term in the sum is associated with a hyperplane through the points $(\lambda_{j_1}, \mu_{j_1}, \ldots, \tau_{j_1}), \ldots, (\lambda_{j_q}, \mu_{j_q}, \ldots, \tau_{j_q})$ and is, in fact, proportional to the $n+p-q-1$ power of the distance of $(s_1, s_2, \ldots, s_q)$ from this plane.

Returning now to the original problem and its notation, we see that $f(r_1, r_2, \ldots, r_q)$ is given by

$$\frac{\Gamma(n+p)}{\Gamma(n+p-q)} \sum_{(j_1,\ldots,j_q) \epsilon R} \cdots \sum \frac{\begin{vmatrix} 1 & r_1 & r_2 & \cdots & r_q \\ 1 & \lambda_{j_1} & \mu_{j_1} & \cdots & \tau_{j_1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \lambda_{j_q} & \mu_{j_q} & \cdots & \tau_{j_q} \end{vmatrix}^{n+p-q-1} \operatorname{sgn} \begin{vmatrix} 1 & \lambda & \mu & \cdots & \tau \\ 1 & \lambda_{j_1} & \mu_{j_1} & \cdots & \tau_{j_1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \lambda_{j_q} & \mu_{j_q} & \cdots & \tau_{j_q} \end{vmatrix}}{\begin{vmatrix} 1 & \lambda & \mu & \cdots & \tau \\ 1 & \lambda_{j_1} & \mu_{j_1} & \cdots & \tau_{j_1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \lambda_{j_q} & \mu_{j_q} & \cdots & \tau_{j_q} \end{vmatrix}^{p} \prod_{j \neq (j_1,\ldots,j_q)} \begin{vmatrix} 1 & \lambda_j & \mu_j & \cdots & \tau_j \\ 1 & \lambda_{j_1} & \mu_{j_1} & \cdots & \tau_{j_1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \lambda_{j_q} & \mu_{j_q} & \cdots & \tau_{j_q} \end{vmatrix}}, \tag{20}$$

where the set R is all $(j_1, \ldots, j_q)$ such that $(r_1, \ldots, r_q)$ is contained in the least convex polyhedron with vertices $(\lambda, \mu, \ldots, \tau), (\lambda_{j_1}, \mu_{j_1}, \ldots, \tau_{j_1}), \ldots, (\lambda_{j_q}, \mu_{j_q}, \ldots, \tau_{j_q})$. In this region, the term in (20) raised to the power p is positive by (19). This is the required distribution. It agrees with

Quenouille's distribution when $p = 0$; this is the case when N is odd. When N is even, $p = \frac{1}{2}$. Furthermore, when $q = 1$, it reduces to Anderson's (1942) distribution for all N.

Quenouille has noted that there are several possible expansions for $f(r_1, r_2, \ldots, r_q)$ when N is odd, i.e. when $p = 0, \lambda = 0, \mu = 0, \ldots, \tau = 0$. In this case the joint density of $r_1, r_2, \ldots, r_q$ is given by

$$\frac{\Gamma(n)}{\Gamma(n-q)} \sum_{(j_1,\ldots,j_q)\,\epsilon\,R} \frac{\begin{vmatrix} 1 & r_1 & r_2 & \ldots & r_q \\ 1 & \lambda_{j_1} & \mu_{j_1} & \ldots & \tau_{j_1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \lambda_{j_q} & \mu_{j_q} & \ldots & \tau_{j_q} \end{vmatrix}^{n-q-1} \operatorname{sgn} \begin{vmatrix} \lambda_{j_1} & \ldots & \tau_{j_1} \\ \vdots & & \vdots \\ \lambda_{j_q} & \ldots & \tau_{j_q} \end{vmatrix}}{\prod_{j+j_1,\ldots,j_q} \begin{vmatrix} 1 & \lambda_j & \mu_j & \ldots & \tau_j \\ 1 & \lambda_{j_1} & \mu_{j_1} & \ldots & \tau_{j_1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \lambda_{j_q} & \mu_{j_q} & \ldots & \tau_{j_q} \end{vmatrix}}, \tag{21}$$

with the summation rule: $(j_1, \ldots, j_q) \,\epsilon\, R(r_1, \ldots, r_q)$ provided $(r_1, r_2, \ldots, r_q)$ falls in the convex closure of $(0, 0, \ldots, 0)$, $(\lambda_{j_1}, \mu_{j_1}, \ldots, \tau_{j_1}), \ldots, (\lambda_{j_q}, \mu_{j_q}, \ldots, \tau_{j_q})$. The domain of the variables $r_1, r_2, \ldots, r_q$ is, of course, the convex closure of $(\lambda_i, \mu_i, \ldots, \tau_i)\,(i = 1, \ldots, n)$. Suppose now that the origin is moved so that we have new variables $r_j^* = r_j + h_j\,(j = 1, \ldots, q)$ with

$$\lambda_i^* = \lambda_i + h_1, \quad \ldots, \quad \tau_i^* = \tau_i + h_q.$$

The joint density of the r_j^* is simply (21) with asterisks on the r's, λ's, μ's, ..., and τ's and the summation rule is found similarly. But this density can be rewritten as

$$\frac{\Gamma(n)}{\Gamma(n-q)} \sum_{(j_1,\ldots,j_q)\,\epsilon\,R^*} \frac{\begin{vmatrix} 1 & r_1 & r_2 & \ldots & r_q \\ 1 & \lambda_{j_1} & \mu_{j_1} & \ldots & \tau_{j_1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \lambda_{j_q} & \mu_{j_q} & \ldots & \tau_{j_q} \end{vmatrix}^{n-q-1} \operatorname{sgn} \begin{vmatrix} \lambda_{j_1}^* & \ldots & \tau_{j_1}^* \\ \vdots & & \vdots \\ \lambda_{j_q}^* & \ldots & \tau_{j_q}^* \end{vmatrix}}{\prod_{j+j_1,\ldots,j_q} \begin{vmatrix} 1 & \lambda_j & \mu_j & \ldots & \tau_j \\ 1 & \lambda_{j_1} & \mu_{j_1} & \ldots & \tau_{j_1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \lambda_{j_q} & \mu_{j_q} & \ldots & \tau_{j_q} \end{vmatrix}}, \tag{22}$$

which is the same as (21) except for the sign term and the summation rule. But the values of the two densities are the same so that, by equating them, identities are obtained. More usefully, however, it may be observed that this is equivalent to using an arbitrary origin to obtain the sign term and the summation rule. Hence, to calculate (21), one may move the origin so that the point $(r_1, r_2, \ldots, r_q)$ is included in the least number of simplexes with vertices, the origin and $(\lambda_{j_1}, \mu_{j_1}, \ldots, \tau_{j_1}), \ldots, (\lambda_{j_q}, \mu_{j_q}, \ldots, \tau_{j_q})$ and so use the shortest possible summation.

4. An approximate density function

In §3, the expression (21) gives the joint density of $r_1, r_2, \ldots, r_q$, where $r_1, r_2, \ldots, r_q$ are defined by (10) in a sample of size $N = 2n+1$. The restriction is $q < n$ or $2q+1 < N$, whereas Quenouille's argument requires $2q+3 < N$. If, then, q is taken as $n-1$, the density is seen

to be independent of the values of $r_1, r_2, \ldots, r_{n-1}$; in fact, it is a constant.† This result leads to an interesting method of approximating the true density function.

Suppose the true polyhedral region of joint variation of $r_1, r_2, \ldots, r_{n-1}$ is replaced by the (convex) region in which the matrix

$$R_{n-1} = \begin{bmatrix} 1 & r_1 & r_2 & \cdots & r_{n-1} \\ r_1 & 1 & r_1 & \cdots & r_{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ r_{n-1} & r_{n-2} & r_{n-3} & \cdots & 1 \end{bmatrix} \tag{23}$$

is positive-definite. It may be verified that this region includes the true polyhedral region and that the vertices of the latter lie on its boundary. The variables $r_{n-1}, \ldots, r_{q+1}$ may be integrated over this region, using a formula from Quenouille,

$$\int \left(\frac{|R_p|}{|R_{p-1}|} \right)^s dr_p = \frac{\Gamma(s+1)}{\Gamma(s+\frac{3}{2})} \pi^{\frac{1}{2}} \left(\frac{|R_{p-1}|}{|R_{p-2}|} \right)^{s+1}, \tag{24}$$

to obtain the approximate density

$$f(r_1, r_2, \ldots, r_q) = \pi^{-\frac{1}{2}q} \prod_{s=n-q}^{n-1} \frac{\Gamma(s+\frac{1}{2})}{\Gamma(s)} \left(\frac{|R_q|}{|R_{q-1}|} \right)^{n-q-1}. \tag{25}$$

An equivalent method is to find the joint density of $r_1, r_{13.2}, \ldots, r_{1n.23} \ldots$. It is easily seen to be proportional to

$$(1 - r_{1n-1.23\ldots}^2)(1 - r_{1n-2.23\ldots}^2)^2 \ldots (1 - r_1^2)^{n-2}. \tag{26}$$

If then $r_1, r_{13.2}, \ldots, r_{1n.23\ldots}$ are all assumed to have the range $(-1, 1)$, they are seen to be independent and the density (25) is obtainable from (26).

The density (25) applies, when $N = 2n+1$, for serial correlations corrected for the mean. Quenouille's conjectured density function, for arbitrary N and serial correlations uncorrected for the mean, is

$$\pi^{-\frac{1}{2}q} \prod_{s=1}^{q} \frac{\Gamma(\frac{1}{2}N - s + 2)}{\Gamma(\frac{1}{2}N - s + \frac{3}{2})} \left(\frac{|R_q|}{|R_{q-1}|} \right)^{\frac{1}{2}(N-2q+1)}. \tag{27}$$

Expressions (25) and (27) do not differ by unity in the value of N as might have been expected. Expression (27) has been criticized by Watson (1951), and Jenkins (1954) has replaced it, for $q = 2$, by a new expression which gives correct moments up to order n. Far less is known about the case when mean corrections are made. However, Jenkins's work for $q = 2$ leads us to the conclusion that (25) gives incorrect moments and is therefore unacceptable.‡

† It is of interest to note that this result may be deduced directly, by elementary methods, from the constancy of the density of the gaps, made by $n-1$ random points on the unit interval.

‡ Of course, comparison of (25) or (26) with Daniels's results show explicitly where they are inadequate. His results apply only when q is $O(1)$ with respect to N. Our results are exact when $q = \frac{1}{2}(N-1)$ and may still have some merit when q is $O(N)$.

REFERENCES

ANDERSON, R. L. (1942). Distribution of the serial correlation coefficient. *Ann. Math. Statist.* **13**, 1–13.

DANIELS, H. E. (1956). The approximate distribution of serial correlation coefficients. *Biometrika*, **43**, 169–85.

DIXON, W. J. (1944). Further contributions to the problem of serial correlation. *Ann. Math. Statist.* **15**, 119–44.

HSU, P. L. (1946). On the asymptotic distribution of certain statistics used in testing the independence between successive observations from a normal population. *Ann. Math. Statist.* **17**, 350–4.

JENKINS, G. M. (1954). Tests of hypotheses in the linear autoregressive model. I. *Biometrika*, **41**, 405–19.

JENKINS, G. M. (1956). Tests of hypotheses in the linear autoregressive model. II. *Biometrika*, **43**, 186–99.

KOOPMANS, T. (1942). Serial correlation and quadratic forms in normal variables. *Ann. Math. Statist.* **13**, 14–33.

MADOW, W. G. (1945). Note on the distribution of the serial correlation coefficient. *Ann. Math. Statist.* **16**, 308–10.

PITMAN, E. J. G. (1937). The 'closest' estimates of statistical parameters. *Proc. Camb. Phil. Soc.* **33**, 212–22.

QUENOUILLE, M. H. (1949). The joint distribution of serial correlation coefficients. *Ann. Math. Statist.* **20**, 561–71.

VON NEUMANN, J. (1941). Distribution of the ratio of the mean square successive difference to the variance. *Ann. Math. Statist.* **12**, 367–95.

VON NEUMANN, J. & MORGENSTERN, O. (1947). *Theory of Games and Economic Behaviour*. Princeton University Press.

WATSON, G. S. (1951). Serial correlation in regression analysis. No. 49 Mimeograph Series, Institute of Statistics, University of North Carolina.

THE APPROXIMATE DISTRIBUTION OF SERIAL CORRELATION COEFFICIENTS

By H. E. DANIELS

Statistical Laboratory, University of Cambridge

1. Introduction and summary

Hotelling's suggestion of a 'circular' definition for the serial correlation coefficient was followed by considerable progress in the distribution theory of such modified statistics by R. L. Anderson (1942), Koopmans (1942), Dixon (1944), Madow (1945) and others. The exact distribution is known for the circular coefficient of any lag from an uncorrelated normal process and, more generally, from a circularly modified normal process of autoregressive type. Quenouille (1949) obtained by the same method the exact joint distribution of circular coefficients of different lags.

The exact distributions are complicated, and a simple and accurate approximation to the distribution of the circular coefficient with known mean was found by Dixon (1944) and Rubin (1945) for the uncorrelated normal process. It was extended to the case of a circular Markov process by Leipnik (1947) following a method due to Madow (1945). The approximation depends on the device of smoothing summation over a discrete set of roots by an approximating integration. Quenouille (1949) conjectured a similar approximate form for the joint distribution, but Watson (1951) and Jenkins (1954) showed that the conjectured form could not be correct. Jenkins developed the correct analogous approximation for the joint distribution of coefficients of lags 1 and 2 with known means.

Without circular modifications the distributional theory is difficult and the field is largely unexplored. For testing independence T. W. Anderson (1948) gave an approximate table of significance points for non-circular lag 1 coefficients with known and fitted means. Watson & Durbin (1951) introduced modified non-circular definitions of the coefficients which have R. L. Anderson's distribution in the uncorrelated case. The case of an unmodified autoregressive process has not been much discussed, though a method due to Bartlett (1953, 1954) is available for obtaining approximate confidence intervals for the parameters, and Quenouille's (1947) approximate goodness of fit tests should be noted.

In the present paper an approach based on the method of steepest descents is adopted to derive the known approximate distributions and to generalize them.* (For an account of the method see, for example, Jeffreys & Jeffreys (1956), Daniels (1954).) The analogue of Leipnik's approximation is found for the distribution of an unmodified coefficient of lag 1, both with known and fitted mean, when the process is of unmodified Markov type. The approximate joint distribution of m successive partial serial correlations is found for an autoregressive process of the mth order, circular modifications being used. The work on the unmodified Markov process could be extended to the general case but we have not done this.

* I have since learnt that Jenkins (1956) has independently obtained many of the results given here by consideration of the moments of smoothed distributions. Of our inevitable differences in notation, the one most likely to confuse is the difference in sign of the autoregression coefficients α_j. My usage, which is that of Bartlett and Quenouille, is adopted for reasons of symmetry.

2. The saddlepoint approximation

In this section some general theory is developed which is fundamental to the rest of the work. We are interested in the distribution of statistics of the form $r = c/c_0$, where c_0 is non-negative. If c_0, c have a joint probability density $f(c_0, c)$ the density for r is

$$h(r) = \int_0^\infty c_0 f(c_0, rc_0)\, dc_0. \tag{2·1}$$

Let

$$M(T_0, T) = E\, e^{T_0 c_0 + Tc}$$

be the joint moment-generating function for c_0 and c. We are concerned only with cases where $M(T_0, T)$ exists in strips of non-zero width containing the imaginary axes in the T_0 and T planes. The usual Fourier inversion formula is most conveniently written as

$$f(c_0, c) = \frac{1}{(2\pi i)^2} \iint M(T_0, T)\, e^{-T_0 c_0 - Tc}\, dT_0 dT, \tag{2·2}$$

the integration being taken along the imaginary axes of T_0 and T, or any allowable deformations of these paths. In particular,

$$\begin{aligned} f(c_0, rc_0) &= \frac{1}{(2\pi i)^2} \iint M(T_0, T)\, e^{-(T_0 + rT)c_0}\, dT_0 dT \\ &= \frac{1}{(2\pi i)^2} \iint M(u - rT, T)\, e^{-uc_0}\, du\, dT, \end{aligned} \tag{2·3}$$

where the integration of $u = T_0 + rT$ is taken over a similar path in the u plane. Inversion of the transform with respect to u gives

$$\int_0^\infty f(c_0, rc_0)\, e^{uc_0}\, dc_0 = \frac{1}{2\pi i} \int M(u - rT, T)\, dT, \tag{2·4}$$

so that, when differentiation is permissible,

$$\int_0^\infty c_0 f(c_0, rc_0)\, e^{uc_0}\, dc_0 = \frac{1}{2\pi i} \int \frac{\partial M(u - rT, T)}{\partial u}\, dT$$

and

$$h(r) = \frac{1}{2\pi i} \int \left. \frac{\partial M(u - rT, T)}{\partial u} \right|_{u=0} dT. \tag{2·5}$$

This is a form of Geary's (1944) extension of Cramér's theorem.

However, we often want to transform T to some other variable z. In (2·4) put

$$T = T(z, u), \tag{2·6}$$

where $T(z, 0)$ maps the T plane on to some region in the z plane, and $\partial T/\partial z$ does not vanish anywhere on the contour except possibly at its termini. Proceeding as before, we find

$$h(r) = \frac{1}{2\pi i} \int \frac{\partial}{\partial u} \left\{ M(u - rT, T) \frac{\partial T}{\partial z} \right\} \Big|_{u=0} dz, \tag{2·7}$$

integration being along the transformed contour in the z plane. This is the form used in the sequel, but the alternative form

$$h(r) = \frac{1}{2\pi i} \int \left\{ \left(\frac{\partial M}{\partial u} \right)_z \frac{\partial T}{\partial z} - \left(\frac{\partial M}{\partial z} \right)_u \frac{\partial T}{\partial u} \right\} \Big|_{u=0} dz \tag{2·8}$$

is sometimes easier to work with in other applications.

Our main task is the approximate evaluation of integrals of type (2·7) when the statistics c_0, c are calculated from a moderately large sample. In the cases considered it is found that the integrand can be written as $\phi(z)[\psi(z)]^n$, where n is the sample size, and the method of steepest descents can be applied. The contour is chosen to pass through a suitable saddlepoint $\hat{z}$ at which $\psi'(\hat{z}) = 0$. It is taken to be a curve of steepest descent, upon which $|\psi(z)|$ decreases most rapidly on either side of $\hat{z}$, and this is one of the branches of

$$\arg \psi(z) = \arg \psi(\hat{z})$$

(see Jeffreys & Jeffreys, 1956). In all the applications considered here the appropriate $\hat{z}$ is found to be real, the corresponding contour of steepest descent through it intersects the real axis normally, and there is no other saddlepoint on the contour so that $|\psi(z)|$ decreases steadily on either side of $\hat{z}$. As n becomes large the major contribution to the integral evidently arises from a small neighbourhood around $\hat{z}$.

The usual procedure is now to expand the integrand in a particular manner about $\hat{z}$, and by termwise integration obtain an asymptotic expansion in powers of n^{-1} whose dominant term is called the saddlepoint approximation. In that way we would find

$$\frac{1}{2\pi i}\int \phi(z)[\psi(z)]^n\,dz \sim \left(\frac{\psi(\hat{z})}{2\pi n \psi''(\hat{z})}\right)^{\frac{1}{2}} \phi(\hat{z})[\psi(\hat{z})]^n, \tag{2·9}$$

the relative error committed being $O(n^{-1})$. But in our applications $\psi(z)$ is such that either the integration can be effected exactly at this stage, or at worst only $\phi(z)$ need be expanded. The dominant term is still the saddlepoint approximation, except that $(2\pi n)^{-\frac{1}{2}}$ is replaced by a slightly different constant.

In discussing the joint distribution of the statistics $r_s = c_s/c_0\,(s = 1, 2, \ldots, m)$, we make use of an extension of (2·7) established in exactly the same way. Let

$$M(T_0, T_1, \ldots, T_m) = E\, e^{T_0 c_0 + T_1 c_1 + \ldots + T_m c_m}.$$

Write $u = T_0 + r_1 T_1 + r_2 T_2 + \ldots + r_m T_m$ and let

$$T_s = T_s(z_1, z_2, \ldots, z_m, u) \quad (s = 1, 2, \ldots, m),$$

be a suitable transformation. Then the joint probability density of $r_1, r_2, \ldots, r_m$ is

$$h(r_1, \ldots, r_m) = \frac{1}{(2\pi i)^m}\int\ldots\int \frac{\partial^m}{\partial u^m}\left\{M\,\frac{\partial(T_1, \ldots, T_m)}{\partial(z_1, \ldots, z_m)}\right\}\Bigg|_{u=0} dz_1 \ldots dz_m. \tag{2·10}$$

The integral can again be approximately evaluated by choosing contours of steepest descent in the planes of $z_1, \ldots, z_m$ passing through saddlepoints, $\hat{z}_1, \ldots, \hat{z}_m$. The details of the work are most easily studied in the actual application of §§ 9–11.

3. Circular Markov process. Known mean

We first discuss the distribution of the circularly defined lag 1 serial correlation coefficient when the process is of circular Markov type with normal residuals and zero means. Let $x_1, x_2, \ldots, x_n$ be such that

$$x_s - \rho x_{s-1} = \epsilon_s, \quad x_0 \equiv x_n \quad (s = 1, 2, \ldots, n),$$

where $\epsilon_1, \ldots, \epsilon_n$ are independent $N(0, 1)$ variables. The joint distribution of the x's is

$$dF = \frac{(1-\rho^n)}{(2\pi)^{\frac{1}{2}n}} \exp\left[-\tfrac{1}{2}\{(1+\rho^2)(x_1^2 + \ldots + x_n^2) - 2\rho(x_1 x_2 + x_2 x_3 + \ldots + x_{n-1}x_n + x_n x_1)\}\right] dx_1 \ldots dx_n. \tag{3·1}$$

The sample estimate of ρ is taken to be $r = c/c_0$, where

$$c_0 = x_1^2 + \ldots + x_n^2, \quad c = x_1 x_2 + \ldots + x_{n-1} x_n + x_n x_1.$$

The moment-generating function for c_0, c is

$$M(T_0, T) = E\, e^{T_0 c_0 + Tc} = (1-\rho^n)\,|\mathbf{A}|^{-\frac{1}{2}},$$

where

$$\mathbf{A} = \begin{bmatrix} 1+\rho^2-2T_0 & -(\rho+T) & 0 & 0 & \ldots & 0 & 0 & -(\rho+T) \\ -(\rho+T) & 1+\rho^2-2T_0 & -(\rho+T) & 0 & \ldots & 0 & 0 & 0 \\ 0 & -(\rho+T) & 1+\rho^2-2T_0 & -(\rho+T) & \ldots & 0 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & \ldots & -(\rho+T) & 1+\rho^2-2T_0 & -(\rho+T) \\ -(\rho+T) & 0 & 0 & 0 & \ldots & 0 & -(\rho+T) & 1+\rho^2-2T_0 \end{bmatrix}.$$

Evaluating its determinant as a circulant we find

$$|\mathbf{A}| = (\rho+T)^n \frac{(1-z^n)^2}{z^n}, \quad z + \frac{1}{z} = \frac{1+\rho^2-2T_0}{\rho+T}, \tag{3·2}$$

and so, writing

$$u = T_0 + rT,$$

$$M(u - rT, T) = \frac{(1-\rho^n)}{(1-z^n)} \frac{(1-2rz+z^2)^{\frac{1}{2}n}}{(1-2\rho r+\rho^2-2u)^{\frac{1}{2}n}}, \tag{3·3}$$

where

$$\rho + T = \frac{z(1-2\rho r+\rho^2-2u)}{(1-2rz+z^2)}. \tag{3·4}$$

The results of § 2 can now be used. Taking $T = T(z, u)$ of (2·6) to be defined by (3·4) we have

$$\left(\frac{\partial T}{\partial z}\right)_u = \frac{(1-z^2)(1-2\rho r+\rho^2-2u)}{(1-2rz+z^2)^2},$$

$$M\left(\frac{\partial T}{\partial z}\right)_u = \frac{(1-\rho^n)(1-z^2)(1-2rz+z^2)^{\frac{1}{2}n-2}}{(1-z^n)(1-2\rho r+\rho^2-2u)^{\frac{1}{2}n-1}},$$

and (2·7) becomes

$$h(r) = \frac{(n-2)(1-\rho^n)}{2\pi i(1-2\rho r+\rho^2)^{\frac{1}{2}n}} \int \frac{(1-z^2)(1-2rz+z^2)^{\frac{1}{2}n-2}}{(1-z^n)}\, dz, \tag{3·5}$$

with

$$\rho + T = \frac{z(1-2\rho r+\rho^2)}{(1-2rz+z^2)}. \tag{3·6}$$

On examining the successive transformations

$$z + \frac{1}{z} = \zeta, \quad \zeta - r = \frac{1-2\rho r+\rho^2}{2(\rho+T)},$$

which together form (3·6), it will be seen that the region $|z| \leqslant 1$ is mapped on to the whole T plane cut along the parts of the real axis exterior to the interval

$$\{-(1+\rho)^2/2(1+r), \quad (1-\rho)^2/2(1-r)\}.$$

Any path in the T plane running from $\tau - i\infty$ through the gap in the real axis to $\tau' + i\infty$ corresponds to a path in $|z| \leqslant 1$ running from $e^{-i\theta}$ to $e^{i\theta}$, where $r = \cos\theta$. The path of integration for (3·5) is therefore of this form.

Consider the factor $(1-2rz+z^2)^{\frac{1}{2}n-2} = \{1-r^2-(z-r)^2\}^{\frac{1}{2}n-2}$ in the integrand. It has a real saddlepoint at $\hat{z}=r$, through which the path of steepest descent is the branch of $\arg(1-2rz+z^2)=0$ crossing the real axis orthogonally. This is just the straight line joining points $e^{-i\theta}$, $e^{i\theta}$, and we choose it to be the path of integration for (3·5).

So far (3·5) is exact, but if we are prepared to tolerate errors which are 'exponentially small', i.e. of magnitude $O(\lambda^n)$ for some $|\lambda|<1$, the factor $1/(1-z^n)$ may be ignored provided neither r nor ρ is near ± 1. This assertion is discussed in the next section; assuming it to be true and putting $z = r+iw(1-r^2)^{\frac{1}{2}}$ $(-1\leqslant w\leqslant 1)$, we find that, ignoring ρ^n,

$$h(r) \sim \frac{(n-2)(1-r^2)^{\frac{1}{2}(n-1)}}{2\pi(1-2\rho r+\rho^2)^{\frac{1}{2}n}} \int_{-1}^{1} (1+w^2)(1-w^2)^{\frac{1}{2}n-2}\,dw$$

$$= \frac{\Gamma(\frac{1}{2}n+1)}{\pi^{\frac{1}{2}}\Gamma(\frac{1}{2}n+\frac{1}{2})} \frac{(1-r^2)^{\frac{1}{2}(n-1)}}{(1-2\rho r+\rho^2)^{\frac{1}{2}n}}, \tag{3·7}$$

which is Leipnik's approximation. The smoothing procedure is thus equivalent to ignoring the factor $1/(1-z^n)$. If this factor is expanded in powers of z^n we obtain on integration the series

$$h(r) = \frac{\Gamma(\frac{1}{2}n+1)(1-\rho^n)}{\pi^{\frac{1}{2}}(1-2\rho r+\rho^2)^{\frac{1}{2}n}} \left\{ \frac{(1-r^2)^{\frac{1}{2}(n-1)}}{\Gamma(\frac{1}{2}n+\frac{1}{2})} - \frac{3}{2^n\Gamma(\frac{3}{2}n+\frac{1}{2})} \frac{d^n}{dr^n}(1-r^2)^{\frac{1}{2}(3n-1)} \right.$$

$$\left. + \frac{5}{2^{2n}\Gamma(\frac{5}{2}n+\frac{1}{2})} \frac{d^{2n}}{dr^{2n}}(1-r^2)^{\frac{1}{2}(5n-1)} - \ldots \right\}, \tag{3·8}$$

which confirms Dixon's observation that when $\rho=0$ the moments of Leipnik's approximation about the origin agree with those of the exact distribution for all orders up to n, since the contributions from terms after the first reduce to zero on partial integration.

4. The error of the approximation

Here we determine an upper bound to the error incurred by the approximation. The effect of ignoring the factor $1/(1-z^n)$ is to alter the integral in (3·5) by an amount

$$\delta = \int \frac{z^n(1-z^2)(1-2rz+z^2)^{\frac{1}{2}n-2}}{(1-z^n)}\,dz,$$

where $z = r+iw(1-r^2)^{\frac{1}{2}}$ $(-1\leqslant w\leqslant 1)$. Since

$$|z|^2 = r^2+w^2(1-r^2),$$

$$|1-z^2| = \{(1-r^2)[1-r+w^2(1+r)][1+r+w^2(1-r)]\}^{\frac{1}{2}} \leqslant 2(1-r^2)^{\frac{1}{2}}, \quad |1-z^n| \geqslant 1-|z|^n,$$

we have

$$|\delta| \leqslant 4(1-r^2)^{\frac{1}{2}n-1} \int_0^1 \frac{[r^2+w^2(1-r^2)]^{\frac{1}{2}n}}{1-[r^2+w^2(1-r^2)]^{\frac{1}{2}n}} (1-w^2)^{\frac{1}{2}n-2}\,dw.$$

The first factor in the integrand keeps it small except near $w=\pm 1$, where the second factor takes over since $1-[r^2+w^2(1-r^2)]^{\frac{1}{2}n} \sim \frac{1}{2}n(1-r^2)(1-w^2)$ only. Let $\int_0^1 = \int_0^k + \int_k^1$, where $0<k<1$. Then

$$\int_0^k < \frac{[r^2+k^2(1-r^2)]^{\frac{1}{2}n}}{1-[r^2+k^2(1-r^2)]^{\frac{1}{2}n}} \int_0^k (1-w^2)^{\frac{1}{2}n-2}\,dw$$

$$= \frac{[r^2+k^2(1-r^2)]^{\frac{1}{2}n}}{1-[r^2+k^2(1-r^2)]^{\frac{1}{2}n}} \tfrac{1}{2}B(\tfrac{1}{2}n-1,\tfrac{1}{2})\{1-I_{1-k^2}(\tfrac{1}{2}n-1,\tfrac{1}{2})\}$$

in the usual notation for complete and incomplete Beta functions. In $k \leqslant w \leqslant 1$,

$$1-[r^2+w^2(1-r^2)]^{\frac{1}{2}n} \geqslant \{1-[r^2+k^2(1-r^2)]^{\frac{1}{2}n}\}\frac{(1-w^2)}{(1-k^2)},$$

since it is a concave function of w^2, so that

$$\int_k^1 \leqslant \frac{(1-k^2)}{1-[r^2+k^2(1-r^2)]^{\frac{1}{2}n}} \int_k^1 (1-w^2)^{\frac{1}{2}n-3}\,dw$$

$$= \frac{(1-k^2)}{1-[r^2+k^2(1-r^2)]^{\frac{1}{2}n}} \tfrac{1}{2}B(\tfrac{1}{2}n-1, \tfrac{1}{2})\, I_{1-k^2}(\tfrac{1}{2}n-2, \tfrac{1}{2}).$$

Hence

$$|\delta| \leqslant \frac{2(1-r^2)^{\frac{1}{2}n-1}}{1-[r^2+k^2(1-r^2)]^{\frac{1}{2}n}} B(\tfrac{1}{2}n-1, \tfrac{1}{2}) \Big\{[r^2+k^2(1-r^2)]^{\frac{1}{2}n}\,[1-I_{1-k^2}(\tfrac{1}{2}n-1, \tfrac{1}{2})]$$
$$+\frac{(n-3)}{(n-4)}(1-k^2)\, I_{1-k^2}(\tfrac{1}{2}n-2, \tfrac{1}{2})\Big\}.$$

For each value of r there is a best value of k which has to be determined numerically. This gives an upper bound for $|\delta|$ which when multiplied by $(n-2)/2\pi(1-2\rho r+\rho^2)^{\frac{1}{2}n}$ gives an upper bound $\Delta(r)$ to the absolute error in Leipnik's approximation. (More strictly the factor $1-\rho^n$ should also be included when ρ^n is not negligible.)

Table 1. *Leipnik's approximate density* $h(r)$. *Error* $\leqslant \Delta(r)$. $n = 20$

$\rho = 0$			$\rho = 0{\cdot}5$					
r	$h(r)$	$\Delta(r)$	r	$h(r)$	$\Delta(r)$	r	$h(r)$	$\Delta(r)$
0	1·8066	0·0050	−1·0	—	—	0·4	1·7511	0·0121
0·1	1·6421	0·0048	−0·9	—	—	0·5	2·0860	0·0254
0·2	1·2258	0·0042	−0·8	—	—			
0·3	0·7375	0·0034	−0·7	—	—	0·55	2·0870	0·0358
0·4	0·3448	0·0024	−0·6	0·0001	—	0·60	1·9339	0·0490
0·5	0·1175	0·0014	−0·5	0·0004	—	0·65	1·6221	0·0640
0·6	0·0260	0·0007	−0·4	0·0023	—	0·70	1·1889	0·0776
0·7	0·0030	0·0002	−0·3	0·0092	—	0·75	0·7185	0·0832
0·8	0·0001	—	−0·2	0·0298	0·0001	0·80	0·3233	0·0725
0·9	—	—	−0·1	0·0817	0·0002	0·85	0·0884	0·0431
1·0	—	—	0	0·1940	0·0005	0·90	0·0092	0·0118
			0·1	0·4059	0·0012	0·95	0·0001	0·0039
			0·2	0·7526	0·0026	1·00	—	—
			0·3	1·2317	0·0056			

Calculations for the case $n = 20$, $\rho = 0$ and 0·5, are shown in Table 1. When $\rho = 0$ the error is seen to be quite negligible in the tails of the distribution, but when $\rho = 0{\cdot}5$ there is a possibility that the upper tail may be materially affected at, say, the 1% level, since $\Delta(r)$ is in that region comparable in magnitude with the ordinates. The error may of course be less than $\Delta(r)$ but cannot be ignored without a more thorough investigation.

5. Circular Markov process. Unknown mean

When the true mean is unknown a suitable estimate of ρ is $r = C/C_0$, where

$$\left.\begin{aligned} C_0 &= (x_1-\bar{x})^2+(x_2-\bar{x})^2+\ldots+(x_n-\bar{x})^2 = c_0-n\bar{x}^2, \\ C &= (x_1-\bar{x})(x_2-\bar{x})+\ldots+(x_{n-1}-\bar{x})(x_n-\bar{x})+(x_n-\bar{x})(x_1-\bar{x}) = c-n\bar{x}^2, \end{aligned}\right\} \tag{5·1}$$

and $\bar{x} = (x_1+x_2+\ldots+x_n)/n$. The moment-generating function for C_0, C is

$$M(T_0, T) = E\, e^{T_0 C_0 + TC} = (1-\rho^n)\left|\mathbf{A}+\frac{2}{n}(T_0+T)\mathbf{1}\mathbf{1}'\right|^{-\frac{1}{2}},$$

with $\mathbf{A}$ as previously defined, and $\mathbf{1}' = [1, 1, \ldots, 1]$ so that $\mathbf{1}\mathbf{1}'$ is the $n \times n$ matrix with unit elements. Since every row of $\mathbf{A}$ sums to $(1-\rho)^2 - 2(T_0+T)$ the determinant can be written as

$$\left|\mathbf{A}+\frac{2}{n}(T_0+T)\mathbf{1}\mathbf{1}'\right| = |\,\mathbf{A}\,|\,|\,\mathbf{I}+\lambda\mathbf{1}\mathbf{1}'\,|, \quad \lambda = \frac{2}{n}\frac{(T_0+T)}{(1-\rho)^2-2(T_0+T)}.$$

The second factor is

$$\begin{vmatrix} 1 & \mathbf{1}' \\ \mathbf{0} & \mathbf{I}+\lambda\mathbf{1}\mathbf{1}' \end{vmatrix} = \begin{vmatrix} 1 & \mathbf{1}' \\ -\lambda\mathbf{1} & \mathbf{I} \end{vmatrix} = \begin{vmatrix} 1+n\lambda & \mathbf{1}' \\ \mathbf{0} & \mathbf{I} \end{vmatrix} = 1+n\lambda,$$

so that $$\left|\mathbf{A}+\frac{2}{n}(T_0+T)\mathbf{1}\mathbf{1}'\right| = \frac{(1-\rho)^2}{(1-\rho)^2-2(T_0+T)}|\,\mathbf{A}\,| = \frac{(\rho+T)^{n-1}}{z^{n-1}}\frac{(1-\rho)^2}{(1-z)^2}(1-z^n)^2,$$

and hence
$$M(u-rT, T) = \frac{(1-\rho^n)}{(1-z^n)}\frac{(1-z)}{(1-\rho)}\frac{(1-2rz+z^2)^{\frac{1}{2}(n-1)}}{(1-2\rho r+\rho^2-2u)^{\frac{1}{2}(n-1)}}, \tag{5·2}$$

with z defined by (3·6). Proceeding as in § 2 we find

$$h(r) = \frac{(n-3)(1-\rho^n)}{2\pi i(1-\rho)(1-2\rho r+\rho^2)^{\frac{1}{2}(n-1)}}\int\frac{(1-z)(1-z^2)(1-2rz+z^2)^{\frac{1}{2}(n-5)}}{(1-z^n)}dz, \tag{5·3}$$

with $z = r+iw(1-r^2)^{\frac{1}{2}}$ $(-1 \leqslant w \leqslant 1)$.

Ignoring the factor $1/(1-z^n)$ again introduces an exponentially small error. Neglecting ρ^n and evaluating the integral we find

$$h(r) \sim \frac{n\Gamma(\frac{1}{2}n-\frac{1}{2})(1-r^2)^{\frac{1}{2}n-1}}{2\pi^{\frac{1}{2}}\Gamma(\frac{1}{2}n)(1-\rho)(1-2\rho r+\rho^2)^{\frac{1}{2}(n-1)}}\left\{(1-r)-\frac{1}{n}(1+r)\right\}. \tag{5·4}$$

As is evident from an expansion of the type (3·8), the moments of r when $\rho = 0$ are exact for all orders up to n. The fact that (5·4) can be negative when $1-r \sim O(n^{-1})$ is not in itself important since approximations of this type in any case break down near $r = 1$. A modified form of the approximation (equation (7·6)) is discussed in § 7.

6. Non-circular Markov process. Known mean

In testing for independence the use of a circularly defined sample serial correlation coefficient may be objectionable on grounds of loss of power and sensitivity to extraneous trends. But when $\rho \neq 0$ the circular definition of the process itself is artificial and can only be justified if the results arrived at by its use are not substantially affected by the assumption of circularity. We now examine the approximate distribution of the sample serial correlation coefficient without introducing circularity assumptions.

The process is defined by

$$x_s - \rho x_{s-1} = \epsilon_s \tag{6·1}$$

for all s, the ϵ's being independent $N(0, 1)$ variables as before. The joint distribution of $x_1, x_2, \ldots, x_n$ is now

$$dF = \frac{(1-\rho^2)^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}n}} \exp[-\tfrac{1}{2}\{x_1^2 + (1+\rho^2)(x_2^2 + \ldots + x_{n-1}^2) + x_n^2 - 2\rho(x_1x_2 + \ldots + x_{n-1}x_n)\}] dx_1 \ldots dx_n. \tag{6·2}$$

The sample estimate of ρ is taken to be $r = c/c_0$, where

$$\left.\begin{aligned} c &= x_1x_2 + x_2x_3 + \ldots + x_{n-1}x_n, \\ c_0 &= \tfrac{1}{2}x_1^2 + x_2^2 + \ldots + x_{n-1}^2 + \tfrac{1}{2}x_n^2. \end{aligned}\right\} \tag{6·3}$$

We have chosen the *intra-class* correlation coefficient between the series $x_1, x_2, \ldots, x_{n-1}$ and $x_2, x_3, \ldots, x_n$. Apart from its intuitive appeal it seems to lead to the simplest analysis both here and in the next section where the mean is fitted. We now have

$$M(T_0, T) = E e^{T_0c_0 + Tc} = (1-\rho^2)^{\frac{1}{2}} |\mathbf{B}|^{-\frac{1}{2}},$$

where

$$\mathbf{B} = \begin{bmatrix} 1-T_0 & -(\rho+T) & 0 & 0 & \ldots & 0 & 0 & 0 \\ -(\rho+T) & 1+\rho^2-2T_0 & -(\rho+T) & 0 & \ldots & 0 & 0 & 0 \\ 0 & -(\rho+T) & 1+\rho^2-2T_0 & -(\rho+T) & \ldots & 0 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & \ldots & -(\rho+T) & 1+\rho^2-2T_0 & -(\rho+T) \\ 0 & 0 & 0 & 0 & \ldots & 0 & -(\rho+T) & 1-T_0 \end{bmatrix}.$$

Then $|\mathbf{B}| = G_n - 2(\rho^2 - T_0) G_{n-1} + (\rho^2 - T_0)^2 G_{n-2}$, where G_n is the determinant of a matrix similar to $\mathbf{B}$ except that all its diagonal elements are $1 + \rho^2 - 2T_0$. Also

$$G_n = (1 + \rho^2 - 2T_0) G_{n-1} - (\rho + T)^2 G_{n-2}$$

with $G_0 = 1$, $G_1 = 1 + \rho^2 - 2T_0$. In this way it is found that

$$|\mathbf{B}| = \frac{(\rho+T)^n}{z^n(1-z^2)} \left\{ \left[1 - \frac{(\rho^2 - 2T_0)z}{(\rho+T)}\right]^2 - z^n \left[z - \frac{(\rho^2 - T_0)}{(\rho+T)}\right]^2 \right\}, \tag{6·4}$$

with z as before. It can again be shown that if $1 - r^2$ is not small, omission of the term in z^n within the bracket incurs an exponentially small error in the final approximation. For brevity we omit it at this stage and write

$$|\mathbf{B}| \sim \frac{(\rho+T)^n}{z^n(1-z^2)} \left[1 - \frac{(\rho^2 - 2T_0)z}{(\rho+T)}\right]^2,$$

whence, after some reduction,

$$M(u - rT, T) \sim \frac{(1 - 2rz + z^2)^{\frac{1}{2}n} (1-\rho^2)^{\frac{1}{2}} (1-z^2)^{\frac{1}{2}}}{(1 - 2\rho r + \rho^2 - 2u)^{\frac{1}{2}n-1} [(1-\rho z)(1 - \rho r - (r-\rho)z) - u(1-z^2)]}, \tag{6·5}$$

with z given by (3·4). Using (2·7) we get the result

$$h(r) \sim \frac{n(1-\rho^2)^{\frac{1}{2}}}{2\pi i (1 - 2\rho r + \rho^2)^{\frac{1}{2}n-1}} \int \phi(z) (1 - 2rz + z^2)^{\frac{1}{2}n-2} dz,$$

where $$\phi(z) = \frac{(1-z^2)^{\frac{3}{2}}}{(1-\rho z)(1-\rho r-(r-\rho)z)}\left\{1+\frac{1}{n}\left[\frac{(1-z^2)(1-2\rho r+\rho^2)}{(1-\rho z)(1-\rho r-(r-\rho)z)}-4\right]\right\}$$

and $z = r+iw(1-r^2)^{\frac{1}{2}}$ $(-1 \leqslant w \leqslant 1)$.

The integral cannot be readily evaluated in closed form. We therefore expand $\phi(z)$ as a power series in $z-r$ and integrate. The odd terms vanish and we get the series

$$\frac{1}{2\pi i}\int \phi(z)(1-2rz+z^2)^{\frac{1}{2}n-2}\,dz$$
$$= \frac{\Gamma(\frac{1}{2}n-1)}{2\pi^{\frac{1}{2}}\Gamma(\frac{1}{2}n-\frac{1}{2})}(1-r^2)^{\frac{1}{2}n-\frac{3}{2}}\left\{\phi(r)-\frac{1}{2(n-1)}(1-r^2)\phi''(r)+\frac{(1-r^2)^2\phi^{\mathrm{iv}}(r)}{2^2\,2!\,(n-1)(n+1)}\right.$$
$$\left.+\ldots+\frac{(-)^s(1-r^2)^s\phi^{(2s)}(r)}{2^s s!\,(n-1)(n+1)\ldots(n+2s-3)}+\ldots\right\}. \quad (6\cdot6)$$

Since $\phi(r) = (1-r^2)^{\frac{1}{2}}/(1-\rho r)\{1+O(n^{-1})\}$ a first approximation to $h(r)$ is

$$h(r) \sim \frac{n}{2\pi^{\frac{1}{2}}}\frac{\Gamma(\frac{1}{2}n-1)}{\Gamma(\frac{1}{2}n-\frac{1}{2})}\frac{(1-\rho^2)^{\frac{1}{2}}(1-r^2)^{\frac{1}{2}n-1}}{(1-\rho r)(1-2\rho r+\rho^2)^{\frac{1}{2}n-1}}\{1+O(n^{-1})\}. \quad (6\cdot7)$$

7. Accuracy of the approximation. Renormalization

The remainder in (6·7) is relatively $O(n^{-1})$, though there has also been neglected an exponentially small term which in samples of moderate size (say, n about 15) may well be comparable in magnitude. To study the accuracy of the approximation adequately would require calculation of an upper bound to the error as in § 4. We do not attempt this here, but instead merely assume that n is large enough for the exponentially small term to be negligible and content ourselves with a heuristic discussion of orders of magnitude. While admittedly not entirely satisfactory it does give some insight into the magnitude of the errors involved, and suggests a device for improving the accuracy without much extra trouble.

Observe first that the variance of r is $O(n^{-1})$ so that values of $h(r)$ outside some range on either side of ρ which is $O(n^{-\frac{1}{2}})$ are negligibly small. Thus $r-\rho$ can be regarded as $O(n^{-\frac{1}{2}})$ over the effective range of r.

Consider the first neglected term in the expansion, the one which is $O(n^{-1})$. If r is replaced by ρ the term is thereby altered by an amount $O(n^{-\frac{3}{2}})$, but it is now independent of r and in fact becomes part of the normalizing constant. One may therefore legitimately write

$$h(r) \sim K\frac{(1-\rho^2)^{\frac{1}{2}}(1-r^2)^{\frac{1}{2}n-1}}{(1-\rho r)(1-2\rho r+\rho^2)^{\frac{1}{2}n-1}}\{1+O(n^{-\frac{3}{2}})\} \quad (7\cdot1)$$

over the effective range of r, where K is an adjusted normalizing constant.* Moreover

$$\frac{1-r^2}{1-2\rho r+\rho^2} = 1-\frac{(r-\rho)^2}{(1-\rho^2)}+O[(r-\rho)^3], \quad (7\cdot2)$$

and so
$$\frac{(1-\rho^2)^{\frac{1}{2}}(1-r^2)^{\frac{1}{2}}}{(1-\rho r)} = 1-\frac{(r-\rho)^2}{2(1-\rho^2)^2}+O[(r-\rho)^3]$$
$$= \left\{\frac{1-r^2}{1-2\rho r+\rho^2}\right\}^{1/(2(1-\rho^2))}+O[(r-\rho)^3]. \quad (7\cdot3)$$

* The device of renormalizing a saddlepoint approximation was used by Cox (1948) without consideration of the magnitude of the remainder.

Consequently, to the same order, $h(r)$ may be written in Leipnik's form as

$$h(r) \sim \frac{\Gamma(\frac{1}{2}N+1)}{\pi^{\frac{1}{2}}\Gamma(\frac{1}{2}N+\frac{1}{2})} \frac{(1-r^2)^{\frac{1}{2}(N-1)}}{(1-2\rho r+\rho^2)^{\frac{1}{2}N}}\{1+O(n^{-\frac{3}{2}})\}, \tag{7·4}$$

with

$$N = n-1+\frac{\rho^2}{1-\rho^2}. \tag{7·5}$$

Actually when $\rho = 0$ the relative magnitude of the remainder reduces to $O(n^{-2})$, since r^2 is now $O(n^{-1})$ and the coefficients of successive terms in the expansion of the integral become functions of r^2 only. The distribution when $\rho = 0$ is approximately the same as that of an *intra-class* correlation coefficient with known mean from a sample of n pairs. (The inter-class coefficient has one fewer degree of freedom.) But when $\rho \neq 0$ the distributions are not the same to the order considered.

The same device can be used to simplify the approximation (5·4) for the circular coefficient with fitted mean. If errors of magnitude $O(n^{-\frac{3}{2}})$ are tolerable the term $(1+r)/n$ in the bracket may be ignored and the distribution renormalized to give

$$h(r) \sim \frac{\Gamma(\frac{1}{2}n+\frac{3}{2})}{2\pi^{\frac{1}{2}}\Gamma(\frac{1}{2}n)\,[n(1-\rho)-(1+\rho)]} \frac{(1-r)\,(1-r^2)^{\frac{1}{2}n-1}}{(1-2\rho r+\rho^2)^{\frac{1}{2}(n-1)}}\{1+O(n^{-\frac{3}{2}})\}. \tag{7·6}$$

So to $O(n^{-\frac{3}{2}})$ the distribution of r with fitted mean is the same as the distribution found by Jenkins (1954) for the first partial coefficient with known mean.

8. Non-circular Markov process. Unknown mean

When the mean is unknown we estimate it by $\bar{x} = (\frac{1}{2}x_1+x_2+\ldots+x_{n-1}+\frac{1}{2}x_n)/(n-1)$ and estimate ρ by $r = C/C_0$, where

$$C = (x_1-\bar{x})\,(x_2-\bar{x})+\ldots+(x_{n-1}-\bar{x})\,(x_n-\bar{x}) = c-(n-1)\,\bar{x}^2,$$

$$C_0 = \tfrac{1}{2}(x_1-\bar{x})^2+(x_2-\bar{x})^2+\ldots+(x_{n-1}-\bar{x})^2+\tfrac{1}{2}(x_n-\bar{x})^2 = c_0-(n-1)\,\bar{x}^2,$$

with c, c_0 defined as in § 6. Then

$$M(T_0, T) = (1-\rho^2)^{\frac{1}{2}}\left|\mathbf{B}+\frac{2(T_0+T)}{n-1}\mathbf{m}\mathbf{m}'\right|,$$

where $\mathbf{B}$ is the matrix of § 6, and $\mathbf{m}' = [\frac{1}{2}, 1, 1, \ldots, 1, \frac{1}{2}]$. The determinant can be evaluated as in § 5, with some extra manipulation to allow for the fact that the sum of the first or last row of $\mathbf{B}$ is not exactly half the sum of any other row. We omit the somewhat lengthy details and record the result that

$$M(u-rT, T) \sim \frac{(1-\rho^2)^{\frac{1}{2}}\,(1-z^2)^{\frac{1}{2}}\,(1-z)\,(1-2rz+z^2)^{\frac{1}{2}(n-1)}}{(1-\rho)\,(1-\rho z)\,[1-\rho r-(r-\rho)\,z-u(1-z^2)]\,(1-2\rho r+\rho^2-2u)^{\frac{1}{2}(n-3)}}$$

$$\times\left\{1+\frac{(1+\rho)^2\,(1-2rz+z^2)\,[(z-\rho)\,(1-\rho z)\,(1-r)+u(1-z^2)]}{(n-1)\,(1-z^2)\,(1-2\rho r+\rho^2-2u)\,[1-\rho r-(r-\rho)\,z-u(1-z^2)]}\right\}^{-\frac{1}{2}},$$

where a term in z^{n-1} has been ignored for the reasons stated, and z is given by (3·6).

If we are content with an approximation having remainder relatively $O(n^{-\frac{3}{2}})$ the last factor may be ignored, the dominant term being taken as before in the expansion of the integral for $h(r)$ and subsequently renormalized. Ultimately it is found that

$$h(r) \sim \frac{K(1-\rho^2)^{\frac{1}{2}}\,(1-r^2)^{\frac{1}{2}(n-3)}\,(1-r)}{(1-\rho r)\,(1-2\rho r+\rho^2)^{\frac{1}{2}(n-3)}}\{1+O(n^{-\frac{3}{2}})\},$$

where K is a normalizing constant. Following the method of § 7 we can replace it to the same order of accuracy by

$$h(r) \sim \frac{\Gamma(\frac{1}{2}N+\frac{3}{2})}{2\pi^{\frac{1}{2}}\Gamma(\frac{1}{2}N)\,[N(1-\rho)-(1+\rho)]}\,\frac{(1-r)(1-r^2)^{\frac{1}{2}N-1}}{(1-2\rho r+\rho^2)^{\frac{1}{2}(N-1)}}\{1+O(n^{-\frac{3}{2}})\}$$

with $N = n-1+\rho^2/(1-\rho^2)$.

The effect of employing non-circular definitions is therefore to replace n by N in the approximate distribution, to $O(n^{-\frac{3}{2}})$, whether the mean is known or fitted. But in the latter case, even when $\rho = 0$ the approximate distribution of r is not the same as that of the intra-class correlation coefficient with fitted mean.

9. Circular autoregressive process of order m. Known mean

The discussion is now extended to cover the mth order autoregressive process. For simplicity only the circularly defined process and statistics are considered. The non-circular case could be dealt with by the methods already used, but the labour involved would be considerable.

The process is defined to be

$$x_s+\alpha_1 x_{s-1}+\alpha_2 x_{s-2}+\ldots+\alpha_m x_{s-m} = \epsilon_s, \quad (s = 1, 2, \ldots, n), \quad x_s \equiv x_{n+s} \tag{9·1}$$

where m is $O(1)$ with respect to n and the ϵ's are independent $N(0, 1)$ variables. It may be concisely written as

$$\mathbf{\Lambda x} = \boldsymbol{\epsilon},$$

where $\mathbf{\Lambda}$ is a circulant with first row $(1, 0, 0, \ldots, 0, \alpha_m, \alpha_{m-1}, \ldots, \alpha_2, \alpha_1)$. Its determinant has the value

$$|\,\mathbf{\Lambda}\,| = \prod_{j=1}^{n}(\omega_j^n+\alpha_1\omega_j^{n-1}+\ldots+\alpha_m\omega_j^{n-m}) = \prod_{t=1}^{m}(1-\theta_t^n), \tag{9·2}$$

where $\omega_j = e^{2\pi ij/n}$ and $\theta_1, \ldots, \theta_m$ are the roots of $\theta^m+\alpha_1\theta^{m-1}+\ldots+\alpha_m = 0$.

The joint distribution of the x's is then

$$dF = \frac{1}{(2\pi)^{\frac{1}{2}n}}\prod_{t=1}^{n}(1-\theta_t^n)\,e^{-\frac{1}{2}\mathbf{x'\Lambda'\Lambda x}}\prod_{j=1}^{n}dx_j. \tag{9·3}$$

Note that $\mathbf{\Lambda'\Lambda}$ is a circulant with first row

$$(1+\alpha_1^2+\alpha_2^2+\ldots+\alpha_m^2,\ \alpha_1+\alpha_1\alpha_2+\alpha_2\alpha_3+\ldots+\alpha_{m-1}\alpha_m,\ \alpha_2+\alpha_1\alpha_3+\alpha_2\alpha_4+\ldots+\alpha_{m-2}\alpha_m, \ldots,$$
$$\alpha_{m-1}+\alpha_1\alpha_m,\ \alpha_m,\ 0, \ldots, 0,\ \alpha_m,\ \alpha_{m-1}+\alpha_1\alpha_m, \ldots, \alpha_1+\alpha_1\alpha_2+\ldots+\alpha_{m-1}\alpha_m).$$

We consider the joint distribution of the m circularly defined coefficients $r_s = c_s/c_0$ $(s = 1, 2, \ldots, m)$, where

$$c_s = x_1x_{s+1}+x_2x_{s+2}+\ldots+x_{n-s}x_n+x_{n-s+1}x_1+\ldots+x_nx_s.$$

The moment-generating function for the c's is

$$M(T_0, T_1, \ldots, T_m) = E\,e^{T_0c_0+T_1c_1+\ldots+T_mc_m}$$
$$= \prod_{t=1}^{m}(1-\theta_t^n)\,|\,\mathbf{\Lambda'\Lambda}-\mathbf{\Theta}\,|^{-\frac{1}{2}}, \tag{9·4}$$

$\boldsymbol{\Theta}$ being a circulant with first row $(2T_0, T_1, T_2, \ldots, T_m, 0, \ldots, 0, T_m, T_{m-1}, \ldots, T_1)$. When $m = 1$, $\boldsymbol{\Lambda}'\boldsymbol{\Lambda} - \boldsymbol{\Theta}$ is just $\mathbf{A}$ of § 3 with $\alpha_1 = -\rho$.

Let us introduce new variables $P, a_1, a_2, \ldots, a_m$ related to the elements of $\boldsymbol{\Lambda}'\boldsymbol{\Lambda} - \boldsymbol{\Theta}$ in the following way:

$$\left.\begin{aligned} 1+\alpha_1^2+\alpha_2^2+\ldots+\alpha_m^2-2T_0 &= P(1+a_1^2+a_2^2+\ldots+a_m^2),\\ \alpha_1+\alpha_1\alpha_2+\alpha_2\alpha_3+\ldots+\alpha_{m-1}\alpha_m-T_1 &= P(a_1+a_1a_2+\ldots+a_{m-1}a_m),\\ \alpha_2+\alpha_1\alpha_3+\alpha_2\alpha_4+\ldots+\alpha_{m-2}\alpha_m-T_2 &= P(a_2+a_1a_3+\ldots+a_{m-2}a_m),\\ \ldots\ldots\ldots\

have Jacobians

$$\frac{\partial(T_0, T_1, \ldots, T_m)}{\partial(P, a_1, \ldots, a_m)} = (-)^{m-1} P^m \begin{vmatrix} \frac{1}{2}(1+a_1^2+\ldots+a_m^2) & a_1 & a_2 & \ldots & a_{m-1} & a_m \\ a_1+a_1a_2+\ldots+a_{m-1}a_m & 1+a_2 & a_1+a_3 & \ldots & a_{m-2}+a_m & a_{m-1} \\ a_2+a_1a_3+\ldots+a_{m-2}a_m & a_3 & 1+a_4 & \ldots & a_{m-3} & a_{m-2} \\ \ldots\ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{m-1}+a_1a_m & a_m & 0 & \ldots & 1 & a_1 \\ a_m & 0 & 0 & \ldots & 0 & 1 \end{vmatrix}$$

$$= \tfrac{1}{2}(-)^{m-1} P^m J(\mathbf{a}),$$

where

$$J(\mathbf{a}) = \begin{vmatrix} 1 & a_1 & a_2 & \ldots & a_{m-1} & a_m \\ a_1 & 1+a_2 & a_1+a_3 & \ldots & a_{m-2}+a_m & a_{m-1} \\ a_2 & a_3 & 1+a_4 & \ldots & a_{m-3} & a_{m-2} \\ \ldots\ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{m-1} & a_m & 0 & \ldots & 1 & a_1 \\ a_m & 0 & 0 & \ldots & 0 & 1 \end{vmatrix}$$

and

$$\frac{\partial(P, a_1, \ldots, a_m)}{\partial(u, a_1, \ldots a_m)} = \left(\frac{\partial P}{\partial u}\right)_{a_1, \ldots, a_m} = -\frac{2}{Q(\mathbf{a}, \mathbf{r})}.$$

Hence

$$\frac{\partial(T_0, T_1, \ldots, T_m)}{\partial(u, a_1, \ldots, a_m)} = (-)^m \frac{[Q(\boldsymbol{\alpha}, \mathbf{r}) - 2u]^m}{Q^{m+1}(\mathbf{a}, \mathbf{r})} J(\mathbf{a}), \tag{9·8}$$

and from (2·10) the joint probability density for $r_1, \ldots, r_m$ is

$$h(\mathbf{r}) = \frac{1}{(2\pi i)^m} \frac{\Gamma(\frac{1}{2}n)}{\Gamma(\frac{1}{2}n-m)} \frac{\prod_{t=1}^{m}(1-\theta_t^n)}{Q^{\frac{1}{2}n}(\boldsymbol{\alpha}, \mathbf{r})} \int \ldots \int \frac{Q^{\frac{1}{2}n-m-1}(\mathbf{a}, \mathbf{r}) J(\mathbf{a})\, da_1, \ldots, da_m}{\prod_{t=1}^{m}(1-\phi_t^m)}. \tag{9·9}$$

The paths of integration in the planes of $a_1, \ldots, a_m$ are chosen to be the lines of steepest descent of $Q(\mathbf{a}, \mathbf{r})$ passing through the saddlepoints $\hat{a}_1, \ldots, \hat{a}_m$ which satisfy

$$\partial Q / \partial \hat{a}_j = 0 \quad (j = 1, 2, \ldots, m),$$

i.e.

$$\mathbf{r} + \mathbf{R}_{m-1}\hat{\mathbf{a}} = 0. \tag{9·10}$$

The saddlepoints are thus the usual least-squares estimators of the α_j's. Also

$$Q(\hat{\mathbf{a}}, \mathbf{r}) = 1 + \mathbf{r}'\hat{\mathbf{a}} = 1 - \mathbf{r}'\mathbf{R}_{m-1}^{-1}\mathbf{r} = |\mathbf{R}_m| / |\mathbf{R}_{m-1}|.$$

Since $Q(\hat{\mathbf{a}}, \mathbf{r})$ is real, $Q(\mathbf{a}, \mathbf{r})$ must remain real on the lines of steepest descent, i.e. if $\mathbf{a} = \boldsymbol{\xi} + i\boldsymbol{\eta}$ they must be such that

$$\mathscr{I} Q(\mathbf{a}, \mathbf{r}) = 2\boldsymbol{\eta}'(\mathbf{r} + \mathbf{R}_{m-1}\boldsymbol{\xi}) = 0.$$

The straight lines $\xi_j = \hat{a}_j$ satisfy this condition. On them, $a_j = \hat{a}_j + i\eta_j$ and

$$Q(\mathbf{a}, \mathbf{r}) = \frac{|\mathbf{R}_m|}{|\mathbf{R}_{m-1}|} - \boldsymbol{\eta}'\mathbf{R}_{m-1}\boldsymbol{\eta}, \tag{9·11}$$

which is a decreasing function of each $|\eta_j|$. They are therefore the required paths of integration.

Since $Q(\mathbf{a}, \mathbf{r}) = 0$ at the end-points of the paths (corresponding to $\mathscr{I}(T_j) = \pm\infty$ in the T_j planes), the domain of integration is $\boldsymbol{\eta}'\mathbf{R}_{m-1}\boldsymbol{\eta} \leqslant |\mathbf{R}_m|/|\mathbf{R}_{m-1}|$.

We shall again ignore the factors $1-\theta_t^n$, $1-\phi_t^n$ since it is not difficult to establish that except in critical cases $|\theta_t| < 1, |\phi_t| < 1$ for all t. (This is equivalent as before to a smoothing operation.) The factor $J(\mathbf{a})$ is a polynomial of degree $m+1$ in the η_j's, and if retained would lead to an approximation to $h(\mathbf{r})$ with an exponentially small remainder. But for simplicity we shall replace it by the constant term $J(\hat{\mathbf{a}})$. The resulting approximation to $h(\mathbf{r})$, when renormalized, will be in error by a factor $1+O(n^{-\frac{3}{2}})$. We then find

$$h(\mathbf{r}) \sim \frac{1}{(2\pi)^m} \frac{\Gamma(\frac{1}{2}n)}{\Gamma(\frac{1}{2}n-m)} \frac{J(\hat{\mathbf{a}})}{Q^{\frac{1}{2}n}(\hat{\mathbf{a}}, \mathbf{r})} \int\cdots\int \left\{ \frac{|\mathbf{R}_m|}{|\mathbf{R}_{m-1}|} - \boldsymbol{\eta}'\mathbf{R}_{m-1}\boldsymbol{\eta} \right\}^{\frac{1}{2}n-m-1} d\eta_1 \ldots d\eta_m. \quad (9{\cdot}12)$$

The integral is readily transformed to the Dirichlet form and has the value

$$\pi^{\frac{1}{2}m} \frac{\Gamma(\frac{1}{2}n-m)}{\Gamma(\frac{1}{2}n-\frac{1}{2}m)} \frac{|\mathbf{R}_m|^{\frac{1}{2}(n-m)-1}}{|\mathbf{R}_{m-1}|^{\frac{1}{2}(n-m-1)}}.$$

The factor $J(\hat{\mathbf{a}})$ can be reduced to some extent as follows. Write it in partitioned form as

$$J(\hat{\mathbf{a}}) = \begin{vmatrix} 1 & \hat{\mathbf{a}}' \\ \hat{\mathbf{a}} & \mathbf{D} \end{vmatrix},$$

where $\mathbf{D}$ is a $m \times m$ matrix having elements $d_{jk} = \hat{a}_{k-j} + \hat{a}_{k+j}$, if we define $\hat{a}_0 = 1$ and $\hat{a}_{m \pm j} = 0$ whenever $j > m$. Also (9·10), which is

$$r_j + \sum_{k=1}^{m} r_{|j-k|}\, \hat{a}_k = 0,$$

can be rearranged as

$$\hat{a}_j + \sum_{k=1}^{m} d_{kj} r_k = 0$$

or

$$\hat{\mathbf{a}} + \mathbf{D}'\mathbf{r} = 0. \quad (9{\cdot}13)$$

Hence

$$J(\hat{\mathbf{a}}) = (1 + \mathbf{r}'\hat{\mathbf{a}})\,|\mathbf{D}| = \frac{|\mathbf{R}_m|}{|\mathbf{R}_{m-1}|}\,|\mathbf{D}|,$$

and the approximation to $h(\mathbf{r})$ is

$$h(\mathbf{r}) \sim K \frac{|\mathbf{R}_m|^{\frac{1}{2}(n-m)}}{|\mathbf{R}_{m-1}|^{\frac{1}{2}(n-m+1)}} \frac{|\mathbf{D}|}{Q^{\frac{1}{2}n}(\boldsymbol{\alpha}, \mathbf{r})} \{1 + O(n^{-\frac{3}{2}})\}. \quad (9{\cdot}14)$$

Since $d_{mj} = 0$ $(j \neq m), d_{mm} = 1$, $|\mathbf{D}|$ reduces to an $(m-1)$th order determinant whose elements are functions of the $\hat{a}_j$'s. As will appear in the next section there is no need to express it in terms of the r_j's.

10. The joint distribution of the leading partial serial correlation coefficients

The distribution simplifies considerably when expressed in terms of partial correlations. We use the convenient notation $r_{s.}$ to denote the partial correlation coefficient between x_s and x_{s+j} conditional on fixed $x_{s+1}, \ldots, x_{s+j-1}$, which may be called the jth leading partial serial correlation coefficient. It is a standard result that

$$|\mathbf{R}_m| = (1-r_1^2)^m (1-r_{2.}^2)^{m-1} (1-r_{3.}^2)^{m-2} \ldots (1-r_{m.}^2) \quad (10{\cdot}1)$$

and also that $\hat{a}_m = -r_{m.}$

The first step is to make a transformation from the r_j's to the $\hat{a}_j$'s. From (9·10) and (9·13) we have $0 = \mathbf{r} + \mathbf{R}_{m-1}\hat{\mathbf{a}} \equiv \hat{\mathbf{a}} + \mathbf{D}'\mathbf{r} = \mathbf{q}$ say, so that

$$0 = \frac{\partial \mathbf{q}}{\partial \hat{\mathbf{a}}} + \frac{\partial \mathbf{q}}{\partial \mathbf{r}} \frac{\partial \mathbf{r}}{\partial \hat{\mathbf{a}}} = \mathbf{R}_{m-1} + \mathbf{D}' \frac{\partial \mathbf{r}}{\partial \hat{\mathbf{a}}},$$

and the required Jacobian is

$$\left| \frac{\partial \mathbf{r}}{\partial \hat{\mathbf{a}}} \right| = -\frac{|\mathbf{R}_{m-1}|}{|\mathbf{D}|}. \tag{10·2}$$

Next we transform from the $\hat{a}_j$'s to the $r_{j\cdot}$'s. Let us introduce the notation $\hat{a}_j \equiv \hat{a}_{j,m}$ to exhibit explicitly the fact that a process of order m is being considered. In this notation $\hat{a}_{j,j} = -r_{j\cdot}$ and (9·10) is $r_j + \sum_{i=1}^{m} r_{|i-j|}\hat{a}_{i,m} = 0$. Subtracting the corresponding equation for order $m-1$ we obtain

$$\sum_{i=1}^{m-1} r_{|i-j|}(\hat{a}_{i,m} - \hat{a}_{i,m-1}) + r_{m-j}\hat{a}_{m,m} = 0.$$

Comparison with

$$\sum_{i=1}^{m-1} r_{|i-j|}\hat{a}_{m-i,m-1} + r_{m-j} = 0$$

gives

$$\hat{a}_{i,m} = \hat{a}_{i,m-1} + \hat{a}_{m-i,m-1}\hat{a}_{m,m}. \tag{10·3}$$

Keeping $\hat{a}_{m,m}$ unaltered, we can use (10·3) to transform the remaining variables

$$\hat{a}_{1,m}, \ldots, \hat{a}_{m-2,m}, \hat{a}_{m-1,m}$$

to

$$\hat{a}_{1,m-1}, \hat{a}_{2,m-1}, \ldots, \hat{a}_{m-2,m-1}, \hat{a}_{m-1,m-1}.$$

Repetition of the procedure ultimately reduces the variables to $\hat{a}_{j,j} = -r_{j\cdot}$ as required.

The Jacobian of (10·3) is the $(m-1)$th order determinant

$$H_m = \begin{vmatrix} 1 & 0 & 0 & \ldots & 0 & 0 & \hat{a}_{m,m} \\ 0 & 1 & 0 & \ldots & 0 & \hat{a}_{m,m} & 0 \\ 0 & 0 & 1 & \ldots & \hat{a}_{m,m} & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & \hat{a}_{m,m} & 0 & \ldots & 0 & 1 & 0 \\ \hat{a}_{m,m} & 0 & 0 & \ldots & 0 & 0 & 1 \end{vmatrix},$$

which has a central element $1 + \hat{a}_{m,m}$ when m is even. Its value is

$$\left.\begin{aligned} H_m &= (1 - r_{m\cdot}^2)^{\frac{1}{2}(m-1)}, && m \text{ odd}, \\ &= (1 - r_{m\cdot})(1 - r_{m\cdot}^2)^{\frac{1}{2}m-1}, && m \text{ even}, \end{aligned}\right\} \tag{10·4}$$

and the Jacobian for the ultimate reduction to the required variables is $H_m H_{m-1} \ldots H_3 H_2$. Combining this with (10·2) and using (10·1) we finally obtain from (9·14) the approximate joint probability density of the $r_{j\cdot}$'s:

$$\frac{K}{Q^{\frac{1}{2}n}(\boldsymbol{\alpha}, \mathbf{r})} \prod_{j \text{ odd}} (1 - r_{j\cdot}^2)^{\frac{1}{2}n-1} \prod_{j \text{ even}} (1 - r_{j\cdot})(1 - r_{j\cdot}^2)^{\frac{1}{2}n-1} \{1 + O(n^{-\frac{1}{2}})\}, \tag{10·5}$$

K being a normalizing constant. It is apparently not possible to decompose $Q(\boldsymbol{\alpha}, \mathbf{r})$ into factors depending on the successive partial correlations.

The most important application of the distribution (provided the circularity assumption can be tolerated) is in testing the hypothesis that $\alpha_m = 0$ by means of the statistic $r_{m\cdot}$. When $\alpha_m = 0$, $Q(\boldsymbol{\alpha}, \mathbf{r})$ does not contain r_m, so if m is odd $r_{m\cdot}$ has density proportional to $(1-r_{m\cdot}^2)^{\frac{1}{2}n-1}$ to $O(n^{-\frac{3}{2}})$, the same as the null distribution of r_1, while if m is even its density is proportional to $(1-r_{m\cdot})(1-r_{m\cdot}^2)^{\frac{1}{2}n-1}$ to $O(n^{-\frac{3}{2}})$, agreeing with that of r_2 found by Jenkins (1954).

It will be remembered that if $J(\mathbf{a})$ had not been replaced by $J(\hat{\mathbf{a}})$ in (9·9) the remainder would have been exponentially small. When $m = 2$, $J(\hat{\mathbf{a}})$ is a cubic in $\mathbf{a}-\hat{\mathbf{a}}$ but the odd terms vanish on integration and there is just one extra term. Calculation shows that the term can be absorbed into the normalizing constant so that the approximation has again an exponentially small remainder. But when $m \geqslant 3$ this is no longer the case.

11. Circular autoregressive process of order m. Unknown mean

Finally, the effect of fitting the mean is considered in the manner of § 5. We require the joint distribution of $r_s = C_s/C_0\,(s = 1, 2, \ldots, m)$, where $C_s = c_s - n\bar{x}^2$. The moment-generating function is

$$M(T_0, T_1, \ldots, T_m) = \prod_{t=1}^{m}(1-\theta_t^n)\left|\boldsymbol{\Lambda}'\boldsymbol{\Lambda}-\boldsymbol{\Theta}+\frac{2}{n}(T_0+T_1+\ldots+T_m)\mathbf{II}'\right|^{-\frac{1}{2}}. \tag{11·1}$$

Each row sum of $\boldsymbol{\Lambda}'\boldsymbol{\Lambda}-\boldsymbol{\Theta}$ is

$$(1+\alpha_1+\ldots+\alpha_m)^2 - 2(T_0+T_1+\ldots+T_m) = P(1+a_1+\ldots+a_m)^2$$

from (9·5), and as before the determinant reduces to

$$\frac{(1+\alpha_1+\ldots+\alpha_m)^2}{P(1+a_1+\ldots+a_m)^2}\,|\,\boldsymbol{\Lambda}'\boldsymbol{\Lambda}-\boldsymbol{\Theta}\,|. \tag{11·2}$$

so that

$$M(u-r_1T_1-\ldots-r_mT_m, T_1, \ldots, T_m) = \frac{(1+a_1+\ldots+a_m)}{(1+\alpha_1+\ldots+\alpha_m)}\frac{Q^{\frac{1}{2}(n-1)}(\mathbf{a}, \mathbf{r})}{[Q(\boldsymbol{\alpha}, \mathbf{r})-2u]^{\frac{1}{2}(n-1)}}\prod_{t=1}^{m}\frac{(1-\theta_t^n)}{(1-\phi_t^n)}. \tag{11·3}$$

The effect on $h(\mathbf{r})$ of having fitted the mean is thus to reduce n to $n-1$ in (9·12), and to introduce the extra factor $1+a_1+\ldots+a_m$ into the integrand, with a suitable readjustment of the normalizing constant. For an approximation with error $O(n^{-\frac{3}{2}})$, the factor can again be replaced by $1+\hat{a}_1+\ldots+\hat{a}_m$ and taken outside the integral.

We have therefore only to evaluate $1+\hat{a}_1+\ldots+\hat{a}_m$ in terms of the $r_{j\cdot}$'s (now with fitted means). From (10·3) we have

$$\begin{aligned}1+\hat{a}_{1,m}+\ldots+\hat{a}_{m,m} &= (1+\hat{a}_{1,m-1}+\ldots+\hat{a}_{m-1,m-1})(1+\hat{a}_{m,m})\\ &= \ldots = (1+\hat{a}_{m,m})(1+\hat{a}_{m-1,m-1})\ldots(1+\hat{a}_{1,1})\end{aligned}$$

or, in the present notation,

$$1+\hat{a}_1+\ldots+\hat{a}_m = (1-r_1)(1-r_{2\cdot})\ldots(1-r_{m\cdot}). \tag{11·4}$$

Hence the approximate density for the $r_{j\cdot}$'s with fitted means is, from (10·5),

$$\frac{K}{Q^{\frac{1}{2}(n-1)}(\boldsymbol{\alpha}, \mathbf{r})}\prod_{j\ \text{odd}}(1-r_{j\cdot})(1-r_{j\cdot}^2)^{\frac{1}{2}(n-3)}\prod_{j\ \text{even}}(1-r_{j\cdot})^2(1-r_{j\cdot}^2)^{\frac{1}{2}(n-3)}\{1+O(n^{-\frac{3}{2}})\}, \tag{11·5}$$

where K is a new normalizing constant.

Thus in testing $\alpha_m = 0$ when the mean is unknown, $r_{m\cdot}$ can be taken as approximately distributed with density proportional to

$$(1-r_{m\cdot})(1-r_{m\cdot}^2)^{\frac{1}{2}(n-3)} \quad \text{when } m \text{ is odd,}$$

and

$$(1-r_{m\cdot})^2(1-r_{m\cdot}^2)^{\frac{1}{2}(n-3)} \quad \text{when } m \text{ is even.}$$

Jenkins (1954) gives the values $E(r_{2\cdot}) \sim -2/(n-1)$ and $E(r_{2\cdot}^2) \sim 1/(n-2)$ for the first two moments of $r_{2\cdot}$ with fitted mean, while Dixon gives the results

$$E(1-r_{2\cdot}^2) = (n-1)/n, \quad E(1-r_{2\cdot}^2)^2 = (n-1)(n+1)/n(n+2).$$

Their values for $E(r_{2\cdot}^2)$ differ by an amount $O(n^{-2})$. Since the error factor in our approximation is actually of the form $1+pr/n+qr^3+\ldots$, it is easy to show that its first and second moments are respectively in error by amounts $O(n^{-2})$ and $O(n^{-3})$. We find

$$E(r_{2\cdot}) \sim -2/(n+1) = -2/n+O(n^{-2})$$

which agrees with Jenkins's value to the required order, but

$$E(r_{2\cdot}^2) \sim (n+5)/(n+1)(n+2) \sim 1/n+2/n^2+O(n^{-3})$$

which agrees with Jenkins's result to $O(n^{-3})$ but not with Dixon's. There is a similar disagreement with Dixon's $E(r_{2\cdot}^4)$.

I am indebted to Miss P. A. Johnson for the computation of Table 1.

REFERENCES

ANDERSON, R. L. (1942). Distribution of the serial correlation coefficient. *Ann. Math. Statist.* **13**, 1–13.

ANDERSON, T. W. (1948). On the theory of testing serial correlation. *Skand. AktuarTidskr.* **31**, 88–116.

BARTLETT, M. S. (1953). Approximate confidence intervals. II. *Biometrika*, **40**, 306–17.

BARTLETT, M. S. (1954). Symposium on interval estimation. Discussion on the papers. *J. R. Statist. Soc.* B, **16**, 208–9.

COX, D. R. (1948). A note on the asymptotic distribution of range. *Biometrika*, **35**, 310–15.

DANIELS, H. E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25**, 631–50.

DIXON, W. J. (1944). Further contributions to the problem of serial correlation. *Ann. Math. Statist.* **15**, 119–44.

GEARY, R. C. (1944). Extension of a theorem by Harald Cramér on the frequency distribution of a quotient of two variables. *J. R. Statist. Soc.* **17**, 56–7.

JEFFREYS, H. & JEFFREYS, B. S. (1956). *Methods of Mathematical Physics*, 3rd ed. Cambridge University Press.

JENKINS, G. M. (1954). Tests of hypotheses in the linear autoregressive model. I. *Biometrika*, **41**, 405–19.

JENKINS, G. M. (1956). Tests of hypotheses in the linear autoregressive model. II. *Biometrika*, **43**, 186–99.

KOOPMANS, T. (1942). Serial correlation and quadratic forms in normal variables. *Ann. Math. Statist.* **13**, 14–33.

LEIPNIK, R. B. (1947). Distribution of the serial correlation coefficient in a circularly correlated universe. *Ann. Math. Statist.* **18**, 86–7.

MADOW, W. G. (1945). Note on the distribution of the serial correlation coefficient. *Ann. Math. Statist.* **16**, 308–10.

QUENOUILLE, M. H. (1947). A large sample test for the goodness of fit of autoregressive schemes. *J. R. Statist. Soc.* B, **11**, 68–84.

QUENOUILLE, M. H. (1949). The joint distribution of serial correlation coefficients. *Ann. Math. Statist.* **20**, 561–71.

RUBIN, H. (1945). On the distribution of the serial correlation coefficient. *Ann. Math. Statist.* **16**, 211–15.

WATSON, G. S. (1951). Serial correlation in regression analysis. Ph.D. thesis, University of North Carolina (unpublished).

WATSON, G. S. & DURBIN, J. (1951). Exact tests of serial correlation using non-circular statistics. *Ann. Math. Statist.* **22**, 446–51.

TESTS OF HYPOTHESES IN THE LINEAR AUTOREGRESSIVE MODEL

II. NULL DISTRIBUTIONS FOR HIGHER ORDER SCHEMES: NON-NULL DISTRIBUTIONS

By G. M. JENKINS

*University College London**

1. Introduction

It is known that the likelihood ratio criterion v_k for testing an autoregressive scheme of order $k-1$ (written as A.R. $(k-1)$) against the alternative hypothesis that it is an A.R. (k) is given by the partial serial correlation between x_i and x_{i-k} when the effects of the intermediate variables have been eliminated. In a previous publication (Jenkins, 1954, to be referred to as I), it was shown that the smoothed form of the distribution of v_2 is given by

$$p(v_2) = \frac{\Gamma(\frac{1}{2}n+\frac{1}{2})}{\Gamma(\frac{1}{2})\,\Gamma(\frac{1}{2}n)}\,(1-v_2^2)^{\frac{1}{2}(n-2)}(1-v_2), \tag{1·1}$$

where $v_2 = \dfrac{r_2-r_1^2}{1-r_1^2}$ and r_k is the circular serial correlation of lag k uncorrected for the mean. The distribution was first derived for a random scheme and then shown to have the same form in an A.R. (1) so that it may be used to test an A.R. (1) against an A.R. (2).

In the present paper, it is proposed to extend this to the case where the serial correlations are corrected for the mean and, also, to construct the relevant distributions for testing higher order schemes. It has been found that up to order 4, the distributions when there is no mean correction are alternately given by

$$p(x) = \frac{\Gamma(\frac{1}{2}n+1)}{\Gamma(\frac{1}{2})\,\Gamma(\frac{1}{2}n+\frac{1}{2})}\,(1-x^2)^{\frac{1}{2}(n-1)} \tag{1·2}$$

and (1·1) according as to whether the order of the alternative hypothesis is odd or even. This result has been proved in the general case by Daniels (1956) using the more elegant method of saddlepoint approximation.

When the serial correlations are corrected for the mean, it will be shown that the distributions of the lower order partial serial correlations are alternately given by (1·1) and

$$p(z) = \{B(\tfrac{1}{2}, \tfrac{1}{2}n-\tfrac{1}{2}) + B(\tfrac{3}{2}, \tfrac{1}{2}n-\tfrac{1}{2})\}^{-1}\,(1-z^2)^{\frac{1}{2}(n-3)}(1-z)^2, \tag{1·3}$$

depending on whether the order of the alternative hypothesis is odd or even. This result has also been proved in the general case by Daniels (1956). The distributions given by (1·2), (1·1) and (1·3) have been designated by type α, type β and type γ respectively.

In the latter half of the paper, the exact and smoothed joint distributions of circular serial correlations (with and without mean correction) are derived, and detailed treatment of these distributions given for the Markoff scheme. The Yule scheme will be discussed at a later stage.

* Now at the Royal Aircraft Establishment, Farnborough.

2. The relation between discrimination and estimation

The procedure envisaged in testing specific hypotheses in an A.R. scheme is part of a more general procedure of discrimination for stationary time-series. This involves a decision as to the nature of the model structure between the three families of alternative hypotheses to randomness given by the autoregressive (A.R.), moving average (M.A.) and linear cyclical (L.C.) processes in the manner which is now being advocated by Rudra (1954, 1955). Rudra's procedure, using a modification of Whittaker's periodogram, enables a decision to be made between the L.C. and M.A. or A.R. schemes. If the latter is adopted, it is suggested that tests for the A.R. and M.A. schemes may be run simultaneously using a method previously given by Rudra (1952).

We shall denote the procedure for picking out the 'best' family of hypotheses by Type-discrimination (or briefly, T-discrimination), and that for selecting the optimum order of hypothesis within a particular model type by Order-discrimination (O-discrimination). In this paper, we shall be concerned with a method of O-discrimination in the A.R. scheme applicable to fairly short series.

It is important to point out that discrimination and estimation are complementary aspects of the inference problem for stationary time-series. It is a convenient property of the A.R. scheme that these two procedures may be carried out independently of one another, discrimination preceding estimation. This is not so for the M.A. scheme where the inference problem reduces to a trial-and-error process in which discrimination and estimation cannot be isolated. In the case of the A.R. scheme, we are in disagreement with the view that one fits the scheme and then tests the goodness of fit by means of the methods of Quenouille (1947) or Whittle (1952). It is not possible to fit the scheme until the order k has been determined, and this is precisely what the goodness of fit test is capable of doing. In fact, the tests of Whittle and Quenouille may be expressed in terms of partial serial correlations.

3. The general method of approach

In the distribution theory of serial correlation coefficients, it is necessary to distinguish between two seemingly independent concepts:

(i) The device of circularity of the serial correlations introduced by Hotelling, which makes it possible to derive exact distributions in the first place.

(ii) The Dixon–Koopmans smoothing technique which leads to more tractable distributions.

In the case of random series, it has been shown that there exists high moment agreement (up to order n) between the smoothed and exact circular distributions, with the consequent assumption that the former provide adequate approximations to the latter. When the variables are autocorrelated, the first moments of these distributions do not agree and it has been concluded that the smoothed distributions are no longer satisfactory. It is suggested that this is to lose sight of the initial aim, viz. to approximate to the distributions of the non-circular statistics.

In this paper, it will be shown that there is considerable evidence that not only are the smoothed distributions tractable but also that they provide much better approximations to the distributions of the non-circular statistics than do the exact circular distributions. This provides a partial solution to the controversial problem as to whether one is losing

power by working with circular statistics. It is suggested that one should always calculate non-circular statistics and use the distribution theory of the smoothed circular statistics.

In I, the author derived an *ad hoc* method for constructing $p_s(v_2)$ (the subscript s indicating that the distribution has been smoothed). This method of approach may be generalized in the following manner:

Writing the partial serial correlation between x_i and x_{i-k} in the form

$$v_k = R_k^{1,(k+1)}/R_{k-1},$$

where

$$R_k = \begin{vmatrix} 1 & r_1 & r_2 & \dots & r_k \\ r_1 & 1 & r_1 & \dots & r_{k-1} \\ \vdots & & & & \vdots \\ r_k & r_{k-1} & \dots & \dots & 1 \end{vmatrix}$$

and $R_k^{1,(k+1)}$ is the cofactor of the first element of the $(k+1)$st row of R_k, the method may be considered in four stages:

(1) $v_1 \equiv r_1, v_2, \dots, v_k$ are assumed to be independent as far as the smoothed distributions are concerned.

(2) r_k is expressed as a function of $v_1, v_2, \dots, v_k$.

(3) Using (1), the moments of v_k may be derived by substituting for the known values of the moments of $v_1, v_2, \dots, v_{k-1}$ and r_k.

(4) The joint distribution of $v_1, v_2, \dots, v_k$ may then be transformed to the joint distribution of $r_1, r_2, \dots, r_k$ and the latter shown to have the same moments as the smoothed distribution. It may then be concluded that stage (4) justifies the assumption made in (1).

The method for deriving the moments of the smoothed distribution has been described in detail in I and will be used quite extensively in what follows.

4. Null distributions for $\bar{v}_1$ and $\bar{v}_2$

It is now proposed to extend the work on $p_s(v_2)$ considered in I so as to cover the case of a fitted mean. It was pointed out in § 9 of this publication that the method for v_2 could not be used for $\bar{v}_2$ since:

(1) $p_s(\bar{r}_1)$ was not known.

(2) The assumption of independence of $\bar{r}_1$ and $\bar{v}_2$ was not tenable since it led to incorrect bivariate moments for the constructed distribution $p(\bar{r}_1, \bar{r}_2)$.

It is necessary to revise these statements in the light of recent work. (1) is not true since $p_s(\bar{r}_1)$ may be constructed from a knowledge of the smoothed moments given by Dixon (1944), viz.

$$\mu'_{2k} = \frac{(2k-1)(2k-3)\dots 1}{(n+1)(n+3)\dots(n+2k-1)}, \tag{4·1}$$

$$\mu'_{2k-1} = -\frac{(2k-1)(2k-3)\dots 1}{(n-1)(n+3)\dots(n+2k-1)}. \tag{4·2}$$

Dixon fitted a type 1 curve using the first two moments, but it is obvious that since they satisfy the relationship

$$\mu'_{2k-1} = -\left(\frac{n+1}{n-1}\right)\mu'_{2k},$$

these are the moments of the function

$$p_s(\bar{r}_1) = \frac{\Gamma(\frac{1}{2}n+\frac{1}{2})}{\Gamma(\frac{1}{2})\,\Gamma(\frac{1}{2}n)}(1-\bar{r}_1^2)^{\frac{1}{2}n-1}\left[1-\frac{n+1}{n-1}\bar{r}_1\right]. \tag{4·3}$$

This is, in fact, another form for the expression obtained by writing $\rho = 0$ in equation (5·4) given by Daniels. The extension to the case where $\rho \neq 0$ is treated generally in the next section.

As far as (2) is concerned, it will be shown in this section that the assumption of independence of $\bar{r}_1$ and $\bar{v}_2$ leads to correct moments of the form $\mathscr{E}(\bar{r}_1^{2s}\bar{r}_2^t)$, while those given by $\mathscr{E}(\bar{r}_1^{2s-1}\bar{r}_2^t)$ differ from their smoothed values by amounts which are negligible for practical purposes.

The assumption of independence of $\bar{r}_1$ and $\bar{v}_2$ leads to the following equations:

$$\mathscr{E}(1-\bar{r}_2)^k = \mathscr{E}(1-\bar{v}_2)^k\,\mathscr{E}(1-\bar{r}_1^2)^k. \tag{4·4}$$

Using (4·1) and (4·2), these may be solved successively to give

$$\frac{n}{n-1} = \mathscr{E}(1-\bar{v}_2)\frac{n}{n+1},$$

$$\frac{n(n+3)}{(n-1)(n+1)} = \mathscr{E}(1-\bar{v}_2)^2\frac{n(n+2)}{(n+1)(n+3)},$$

$$\frac{n(n+4)(n+5)}{(n-1)(n+1)(n+3)} = \mathscr{E}(1-\bar{v}_2)^3\frac{n(n+2)(n+4)}{(n+1)(n+3)(n+5)},$$

etc., and these result in the following forms for the lower moments of $\bar{v}_2$:

$$\mathscr{E}(\bar{v}_2) = -\frac{2}{n-1}, \qquad \mathscr{E}(\bar{v}_2^2) = \frac{1}{n+2}\left(\frac{n+3}{n-1}\right),$$

$$\mathscr{E}(\bar{v}_2^3) = -\frac{6}{(n-1)(n+2)}, \quad \mathscr{E}(\bar{v}_2^4) = \frac{3}{(n+2)(n+4)}\left(\frac{n+7}{n-1}\right).$$

It is not difficult to show that the general results take the form

$$\mathscr{E}(\bar{v}_2^{2k-1}) = -\frac{2(2k-1)(2k-3)\ldots 1}{(n-1)(n+2)\ldots(n+2k-2)}, \tag{4·5}$$

$$\mathscr{E}(\bar{v}_2^{2k}) = \left(\frac{n+4k-1}{n-1}\right)\frac{(2k-1)(2k-3)\ldots 1}{(n+2)(n+4)\ldots(n+2k)}, \tag{4·6}$$

and the latter may be rearranged to give

$$\mathscr{E}(\bar{v}_2^{2k}) = \frac{n}{n-1}\left\{\frac{(2k-1)(2k-3)\ldots 1}{n(n+2)\ldots(n+2k-2)} + \frac{(2k+1)(2k-1)\ldots 1}{n(n+2)\ldots(n+2k)} - \frac{2}{n}\left[\frac{(2k-1)(2k-3)\ldots 1}{(n+2)\ldots(n+2k)}\right]\right\}.$$

It follows immediately that these are the moments of the function

$$p(\bar{v}_2) = k(1-\bar{v}_2^2)^{\frac{1}{2}(n-3)}\left\{(1-\bar{v}_2)^2 - \frac{2}{n}(1-\bar{v}_2^2)\right\}, \tag{4·7}$$

where

$$k^{-1} = B(\tfrac{1}{2}, \tfrac{1}{2}n-\tfrac{1}{2}) + B(\tfrac{3}{2}, \tfrac{1}{2}n-\tfrac{1}{2}) - \frac{2}{n}B(\tfrac{1}{2}, \tfrac{1}{2}n+\tfrac{1}{2}).$$

It may be seen that the dominant term agrees with the saddlepoint approximation given by Daniels; it is also a type γ distribution to the first order.

It is necessary now to derive the moments of the constructed distribution obtained by transforming $p(\bar{r}_1, \bar{v}_2) = p(\bar{r}_1)\, p(\bar{v}_2)$. The problem is not entered upon in its greatest generality here, since the method is similar to the one used in I for $p(\bar{v}_2)$. For example, the moment $\bar{\mu}_{21}$ of the constructed distribution may be evaluated as follows:

$$\begin{aligned}\mathscr{E}(\bar{r}_1^2 \bar{r}_2) &= \mathscr{E}(\bar{r}_1^4) + \mathscr{E}(\bar{v}_2)\, \mathscr{E}\{\bar{r}_1^2(1-r_1^2)\} \\ &= \frac{n-3}{(n-1)(n+1)(n+3)},\end{aligned}$$

and this agrees with the value derived for the exact moment (and hence the smoothed moment if its order is less than n) by the author in I.

In a similar manner, it may be shown that

$$\begin{aligned}\mathscr{E}(\bar{r}_1 \bar{r}_2) &= \mathscr{E}(\bar{r}_1^3) + \mathscr{E}(\bar{v}_2)\, \mathscr{E}\{\bar{r}_1(1-\bar{r}_1)^2\} \\ &= -\frac{1}{(n-1)^2}\left\{1 - \frac{3}{n}\right\},\end{aligned}$$

which is to be compared with the value given in I, viz. $-1/(n-1)^2$. Generally, it appears that moments of the form $\bar{\mu}_{2s,t}$ are correct whilst those of the form $\bar{\mu}_{(2s-1),t}$ do not agree with the moments of the smoothed distribution.

It may thus be concluded that as far as the smoothed distributions are concerned, the partial serial correlations are not independent when the serial correlations are corrected for the mean. Also that the type β and type γ distributions in this case are not smoothed distributions but represent dominant terms in these forms.

5. Null distributions for v_3 and v_4

The general method will now be applied to the construction of $p_s(v_3)$ and $p_s(v_4)$. As will become apparent, the situation is much more complicated for these cases, and it seems difficult to extend the analysis to cover the cases where $k > 4$. Since the method of approach has now been described in detail, the algebraic manipulation will be omitted and only the main stages in the argument included.

In terms of the notation of § 3, we may write

$$v_3 = R_3^{1,4}/R_2 = \frac{r_1(r_1^2 - r_2) - r_1 r_2(1-r_2) + r_3(1-r_1^2)}{R_2}.$$

This may then be simplified to give

$$v_3 = \frac{(r_3 - r_1) + r_1(1-r_1^2)(1-v_2)^2}{(1-r_1^2)(1-v_2^2)}, \tag{5·1}$$

and then rewritten to satisfy condition (2) of § 4 in the form

$$r_3 = v_3(1 - T_{20}^2) + r_1 - r_1(1-r_1^2)(1-v_2)^2, \tag{5·2}$$

where

$$(1 - T_{k0}^2) = (1-r_1^2)(1-v_2^2)\ldots(1-v_k^2). \tag{5·3}$$

In a similar manner, it may be shown that

$$v_4 = R_4^{1,5}/R_3 = 1 + \begin{vmatrix} r_4-1 & r_1 & r_2 & r_3 \\ r_3-r_1 & 1 & r_1 & r_2 \\ 0 & r_1 & 1 & r_1 \\ r_1-r_3 & r_2 & r_1 & 1 \end{vmatrix} \div R_3.$$

This may be simplified using (5·2), yielding

$$(v_4-1)(1-T_{30}^2) = (r_4-1)+(r_1-r_3)[r_1(1-v_2)-v_3(1+v_2)], \tag{5·4}$$

and rewritten to satisfy condition (2) of § 3 in the form

$$(v_4-1)(1-T_{30}^2) = (r_4-1)+r_1^2(1-r_1^2)(1-v_2)^3-2r_1v_3(1-r_1^2)(1-v_2^2)(1-v_2) + v_3^2(1-r_1^2)(1-v_2^2)(1+v_2). \tag{5·5}$$

Equations (5·1) and (5·4) emerge as very simple formulae for calculating v_3 and v_4. The saving in labour by using these formulae instead of the original determinantal forms is considerable.

We now proceed to determine the smoothed moments of v_3 using the known results for the moments of v_2 and also that

$$\left.\begin{aligned} \mathscr{E}(r_j^{2i-1}) &= 0, \\ \mathscr{E}(r_j^{2i}) &= \delta = \frac{(2i-1)(2i-3)\dots 1}{(n+2)(n+4)\dots(n+2i)}. \end{aligned}\right\} \tag{5·6}$$

Taking expectations throughout in (5·2) and using condition (1) of § 3, it follows immediately that $\mathscr{E}(v_3) = 0$. Taking the cubes of both sides and then expectations, it follows that $\mathscr{E}(v_3^3) = 0$, since $\mathscr{E}(v_3) = 0$ and the remaining terms involve expectations of odd powers of r_1 and r_3 which are known to be zero from (5·6). Proceeding in an analogous manner, it may be shown by means of an inductive argument that the odd moments are all zero and it remains to find the even moments. These may be derived by the method used in the previous section, viz. by considering (5·2) raised to successive powers. However, the algebra is slightly simpler if we proceed as follows. Since $(r_3-r_1)^2$ may be written in the form

$$v_3^2(1-r_1^2)^2(1-v_2^2)^2+r_1^2(1-r_1^2)^2(1-v_2)^4-2v_3r_1(1-r_1^2)^2(1-v_2^2)(1-v_2)^2,$$

it follows by taking expectations that

$$\mathscr{E}(r_3-r_1)^2 = \frac{n}{n+4}\mathscr{E}(v_3^2)+\frac{n+8}{(n+2)(n+4)}. \tag{5·7}$$

It is possible to show by differentiation of the smoothed characteristic function in the manner described in I that $\mathscr{E}_s(r_1r_3) = 0$. It follows therefore that $\mathscr{E}_s(r_1-r_3)^2 = 2/(n+2)$, and substitution in (5·7) yields $\mathscr{E}_s(v_3^2) = 1/(n+2)$.

In a similar manner, we may write

$$\mathscr{E}(r_3-r_1)^4 = \frac{n}{n+8}\mathscr{E}(v_3^4)-\frac{3(n+14)(n+16)}{(n+2)(n+4)(n+6)}+\frac{6n(n+12)}{(n+2)\dots(n+8)}, \tag{5·8}$$

and the left-hand side may be evaluated using the following smoothed moments which are obtained quite easily:

$$\mathscr{E}_s(r_1 r_3^3) = 0. \tag{5·9}$$

$$\mathscr{E}_s(r_1^2 r_3^2) = \frac{n+12}{(n+2)(n+4)(n+6)}, \tag{5·10}$$

$$\mathscr{E}_s(r_1^3 r_3) = \frac{6}{(n+2)(n+4)(n+6)}. \tag{5·11}$$

Hence $$\mathscr{E}_s(r_3 - r_1)^4 = \frac{12(n+7)}{(n+2)(n+4)(n+6)},$$

and substitution in (5·8) yields $$\mathscr{E}_s(v_3^4) = \frac{3}{(n+2)(n+4)}.$$

Continuing the process, it appears that the moments of v_3 are those of type α form. In fact, this is quite easily verified by writing down the joint distribution of v_1, v_2, v_3 and transforming to $p_s(r_1, r_2, r_3)$ which may then be integrated to give $p_s(r_3)$, showing that the constructed distribution has the correct marginal distributions.

It has not been found possible to treat the problem of comparing the multivariate moments of the smoothed and constructed distributions in its greatest generality. The method will be illustrated in the case of $\mathscr{E}(r_1^{2s-1} r_3)$.

From (5·2) we obtain an expression for the moment of the constructed distribution in the form

$$\begin{aligned}\mathscr{E}(r_1^{2s-1} r_3) &= \mathscr{E}(r_1^{2s}) - \mathscr{E}\{r_1^{2s}(1-r_1^2)\}\,\mathscr{E}(1-v_2)^2 \\ &= \frac{(2s-2)(2s-1)(2s-3)\ldots 1}{(n+2)(n+4)\ldots(n+2s+2)}.\end{aligned} \tag{5·12}$$

The corresponding smoothed moment may be derived from the smoothed moment-generating function which is obtained by writing $\alpha_i = 0$ and $k = 3$ in (6·7), viz.

$$\phi_s(\theta_0, \theta_1, \theta_2, \theta_3) = \exp\left[-\frac{n}{4\pi}\int_0^{2\pi} \log\{1 - \theta_0 - \theta_1 \cos\alpha - \theta_2 \cos 2\alpha - \theta_3 \cos 3\alpha\}\, d\alpha\right].$$

From this we may obtain $\mathscr{E}_s(r_1^{2s-1} r_3)$ by differentiating partially with respect to θ_3, placing $\theta_2 = 0 = \theta_3$ and then expanding this in the form of an infinite series in powers of $x = \theta_1/(1-\theta_0)$ (for further details, see § 7 of I). Thus

$$\left[\frac{\partial \phi_s}{\partial \theta_3}\right]_{\substack{\theta_2=0\\ \theta_3=0}} = \exp\left[-\frac{n}{4\pi}\int_0^{2\pi} \log\{1-\theta_0-\theta_1\cos\alpha\}\, d\alpha\right] \frac{n}{2\pi}\int_0^{2\pi} \frac{\cos 3\alpha\, d\alpha}{(1-\theta_0-\theta_1\cos\alpha)},$$

and the first integral may be expanded in the form of an infinite series and integrated term by term giving

$$2\pi \sum_{j=2}^{\infty} \frac{(2j-2)(2j-1)(2j-3)\ldots 1}{2^{j+1}(j+1)!}\, x^{2j-1}.$$

The other factor may be written as $[\frac{1}{2}\{1+\sqrt{(1-x^2)}\}]$, and a series expansion for this has been given in I. The two series may then be multiplied giving finally

$$\left[\frac{\partial \phi_s}{\partial \theta_3}\right]_{\substack{\theta_2=0\\ \theta_3=0}} = \left[\frac{i}{8}x^3 + \sum_{j=3}^{\infty} \frac{i(i+j+2)(i+j+3)\ldots(i+2j-1)}{2^{2j-2}(j-2)!}\, x^{2j-1}\right] \quad (i = \tfrac{1}{2}n),$$

and by integration it may be shown that $\mathscr{E}_s(r_1^{2s-1}r_3)$ is the same as the expression given by (5·12). In this manner it is possible to show that a large number of the simpler trivariate moments of the constructed distribution agree with those derived from the smoothed characteristic function.

The procedure for constructing $p_s(v_4)$ is similar in principle but leads to much more complicated equations for the moments. It has been verified that these are the moments of the type β distribution.

Fairly extensive sampling experiments have been conducted on artificial series which demonstrate the fact that the distributions of the v_k alternate in the case of random series. It has also been shown that the distribution theory provides a basis for O-discrimination in the A.R. scheme. For example, in the case of schemes of orders 1, 2 and 3 using samples of 22 upwards, it has been shown that the number of correct classifications is high except in the case of short series with very strong autocorrelation. It is conjectured that this effect may be explained by the fact that $\mathscr{E}(v_k)$ in an A.R. $(k-1)$ depends on $\alpha_1, \alpha_2, \ldots, \alpha_{k-1}$ to $O(n^{-2})$. It is not difficult to show that $p_s(v_k \mid \alpha_1 \ldots \alpha_{k-1})$ is the same as the distribution of v_k in a random scheme, so that it appears that the smoothed distributions are inadequate when n is fairly small and $\alpha_1 \ldots \alpha_{k-1}$ are close to their limiting values.

6. Some general theorems on non-null distributions

There seems to be some confusion in the literature about the derivation of non-null distributions, and a number of those put forward are not density functions at all since they do not integrate out to unity.

A fundamental theorem in this work is that due to Madow (1945), but this applies to exact distributions only. In this section it is proposed to apply a unified method in order to derive four distributions, viz. the exact and smoothed non-null joint distributions of circular serial correlations with and without mean correction. In view of its practical importance as an approximation to the exact joint distribution of non-circular statistics, the smoothed distribution with serials corrected for the mean is of particular interest. In the case of the Markoff scheme, it leads to a smoothed distribution for the first lag serial correlation corrected for the mean, analogous to Leipnik's distribution.

Defining the modified standardized variance and serial covariances by $c'_j = \dfrac{1}{2\sigma^2}\sum_{i=1}^{n} x_i x_{i+j}$, with analogous expressions $\bar{c}'_j$ when there is a mean correction, it may be shown that the joint characteristic function of these statistics in the A.R. (k) scheme

$$\alpha_0 x_t + \alpha_1 x_{t-1} + \ldots + \alpha_k x_{t-k} + z_t = 0 \quad (\alpha_0 = -1) \tag{6·1}$$

may be written in the form

$$\phi(\theta_0, \ldots, \theta_k) = |J_k| \prod_{j=1}^{n} \left\{ b_0 + b_1 \cos\frac{2\pi j}{n} + \ldots + b_k \cos\frac{2\pi jk}{n} \right\}^{-\frac{1}{2}}, \tag{6·2}$$

and

$$\bar{\phi}(\theta_0, \ldots, \theta_k) = \frac{|J_k|}{1-\alpha_1 \ldots - \alpha_k} \prod_{j=1}^{n-1} \left\{ b_0 + b_1 \cos\frac{2\pi j}{n} + \ldots + b_k \cos\frac{2\pi jk}{n} \right\}^{-\frac{1}{2}}, \tag{6·3}$$

respectively, where $b_j = a_j - i\theta_j$, $a_j = 2\sum_{i=1}^{k} \alpha_i \alpha_{i+j}$ $(j = 1, 2, \ldots, k)$ and $a_0 = \sum_{i=1}^{k} a_i^2$. $|J_k|$ is the Jacobian of the transformation from the z's to the x's in (6·1) and is given by the circulant

$$|J_k| = \prod_{j=1}^{n} \{1 - \alpha_1 \omega_j - \alpha_2 \omega_j^2 - \ldots - \alpha_k \omega_j^k\}. \tag{6·4}$$

It is necessary now to derive the smoothed forms of these functions. The smoothing process will be defined formally in the following manner:

If

$$\lim_{n\to\infty} -\frac{2}{n}\log\phi(\theta_0,\ldots,\theta_k,\alpha_1,\ldots,\alpha_k,n) = L(\theta_0,\ldots,\theta_k,\alpha_1,\ldots,\alpha_k)$$

then the smoothed characteristic function will be defined by

$$\phi_s(\theta_0,\ldots,\theta_k) = \exp L^{-\frac{1}{2}n}. \tag{6·5}$$

Since this procedure is rather arbitrary, it will be necessary to show that this function satisfies the relationship

$$\phi_s(\theta,0,\ldots,0) = 1. \tag{6·6}$$

It is obvious that it satisfies all the other necessary conditions for being a characteristic function which have been given by Kendall (1948), since (6·2) satisfies these conditions and the latter are unaffected by smoothing. In order that (6·5) is to represent a true characteristic function, it must satisfy the further condition that, when inverted, it yields a positive function.

The following theorem will now be stated and its proof given in outline.

THEOREM. *Provided that* $\alpha_1,\alpha_2,\ldots,\alpha_k$ *are such that* (6·1) *generates a stationary process, the smoothed characteristic function corresponding to* (6·2) *is given by*

$$\phi_s(\theta_0,\ldots,\theta_k) = \exp\left[-\frac{n}{4\pi}\int_0^{2\pi}\log\{b_0+b_1\cos\psi+\ldots+b_k\cos k\psi\}\,d\psi\right], \tag{6·7}$$

and it satisfies condition (6·6).

The proof resolves itself into a consideration of the separate limits of the Jacobian and the finite product in (6·2). Using the following lemma, it may be shown that the contribution due to the Jacobian is zero.

LEMMA. *Provided that the* $\alpha_1,\ldots,\alpha_k$ *are such that* (6·1) *generates a stationary process,* $|J_k|\to 1$ *as* $n\to\infty$.

Proof. This follows by observing that $|J_k|$ may be written in the form

$$(1-A_1^n)(1-A_2^n)\ldots(1-A_k^n), \tag{6·8}$$

where $A_1,A_2,\ldots,A_k$ are the roots of the equation

$$x^k-\alpha_1x^{k-1}-\ldots-\alpha_k = 0.$$

From a general theorem due to Wold (1938), it is known that the roots of this equation have moduli less than unity so that each term in (6·8) tends to 1 as $n\to\infty$.

The proof of the theorem is completed by showing that the product term in (6·2) tends to the expression (6·7), and the latter may be shown to satisfy (6·6) by justifying an interchange of limit operations.

Since

$$\overline{\phi}(\theta_0,\ldots,\theta_k) = \phi(\theta_0,\ldots,\theta_k)\left[\frac{(1-\alpha_1-\alpha_2-\ldots-\alpha_k)^2-i\theta_0-i\theta_1-\ldots-i\theta_k}{(1-\alpha_1-\alpha_2-\ldots-\alpha_k)^2}\right]^{\frac{1}{2}},$$

it follows that the same relationship holds for the corresponding smoothed functions. We now proceed to express the non-null distributions corresponding to these characteristic functions in terms of their null distributions. For example, Fourier inversion of (6·2) yields

$$p(c_0,\ldots,c_k\,|\,\alpha_i) = \frac{1}{(2\pi i)^{k+1}}\int_0^\infty\ldots\int_0^\infty e^{i\Sigma c_j\theta_j}\,\phi(\theta_0,\ldots,\theta_k)\prod_{l=0}^{k}d\theta_l,$$

whilst the transformation $\lambda_j = (a_j - i\theta_j)$ results in

$$p(c_0, \ldots, c_k \mid \alpha_i) = \exp\left[-\sum_{=0}^{k} a_j c_j\right] \mid J_k \mid p(c_0, \ldots, c_k \mid \alpha_i = 0).$$

It is then possible

(i) to transform to new variables $c_0, r_1, \ldots, r_k$,

(ii) to use the fact that c_0 is independently distributed of $p(r_1, r_2, \ldots, r_k)$ when the $\alpha_i = 0$, and

(iii) to integrate with respect to c_0 between the limits 0 and ∞.

In the case of (6·2) and (6·3) this leads to

$$p(r_1, r_2, \ldots, r_k \mid \alpha_i) = \frac{\mid J_k \mid}{[a_0 + a_1 r_1 + \ldots + a_k r_k]^{\frac{1}{2}n}} p(r_1, \ldots, r_k \mid \alpha_i = 0), \tag{6·9}$$

$$p(\bar{r}_1, \bar{r}_2, \ldots, \bar{r}_k \mid \alpha_i) = \frac{\mid J_k \mid (1 - \alpha_1 - \ldots - \alpha_k)^{-1}}{[a_0 + a_1 \bar{r}_1 + \ldots + a_k \bar{r}_k]^{\frac{1}{2}(n-1)}} p(\bar{r}_1, \ldots, \bar{r}_k \mid \alpha_i = 0), \tag{6·10}$$

respectively, where the null hypothesis distributions

$$p(r_1, \ldots, r_k \mid \alpha_i = 0) \quad \text{and} \quad p(\bar{r}_1, \ldots, \bar{r}_k \mid \alpha_i = 0)$$

have been given by Quenouille (1949) and for a more general case by Watson (1956). It is to be noted that (6·9) and (6·10) are different from the distributions for an A.R. (k) scheme given by Quenouille (1949).

In the case of the smoothed distributions, the procedure is identical except that in this case it is necessary to prove that $p_s(c_0) \equiv p(c_0)$ is independent of $p_s(r_1, r_2, \ldots, r_k)$. This follows by a trivial modification of a theorem due to Pitman (1937), replacing the discrete set of characteristic vectors of the quadratic forms in the definition of the serial correlations by a continuous set in the same manner as Koopmans (1942).

The above method then leads to the following smoothed distributions:

$$p_s(r_1, \ldots, r_k \mid \alpha_i) = \frac{1}{[a_0 + a_1 r_1 + \ldots + a_k r_k]^{\frac{1}{2}n}} p_s(r_1, \ldots, r_k \mid \alpha_i = 0), \tag{6·11}$$

$$p_s(\bar{r}_1, \ldots, \bar{r}_k \mid \alpha_i) = \frac{(1 - \alpha_1 - \ldots - \alpha_k)^{-1}}{[a_0 + a_1 \bar{r}_1 + \ldots + a_k \bar{r}_k]^{\frac{1}{2}(n-1)}} p_s(\bar{r}_1, \ldots, \bar{r}_k \mid \alpha_i = 0). \tag{6·12}$$

(6·11) is the generalization of Leipnik's distribution, whilst (6·12) is new and corresponds to equation (9·14) given by Daniels. In the next section we shall be concerned with (6·11) and (6·12) when $k = 1$.

7. NON-NULL DISTRIBUTIONS ON THE MARKOFF SCHEME

In an A.R. (1), equations (6·9), (6·10), (6·11) and (6·12) reduce to

$$p(r_1 \mid \alpha_1) = (1 - \alpha_1^n)\, p(r_1 \mid \alpha_1 = 0)\, [1 + \alpha_1^2 - 2\alpha_1 r_1]^{-\frac{1}{2}n}, \tag{7·1}$$

$$p(\bar{r}_1 \mid \alpha_1) = \left(\frac{1 - \alpha_1^n}{1 - \alpha_1}\right) p(\bar{r}_1 \mid \alpha_1 = 0)\, [1 + \alpha_1^2 - 2\alpha_1 \bar{r}_1]^{-\frac{1}{2}(n-1)}, \tag{7·2}$$

$$p_s(r_1 \mid \alpha_1) = p_s(r_1 \mid \alpha_1 = 0)\, [1 + \alpha_1^2 - 2\alpha_1 r_1]^{-\frac{1}{2}n}, \tag{7·3}$$

$$p_s(\bar{r}_1 \mid \alpha_1) = \frac{p_s(\bar{r}_1 \mid \alpha_1 = 0)}{(1 - \alpha_1)} [1 + \alpha_1^2 - 2\alpha_1 \bar{r}_1]^{-\frac{1}{2}(n-1)}. \tag{7·4}$$

The null hypothesis distributions for (7·1) and (7·2) have been given by R. L. Anderson (1942), but these exact distributions are of no use practically since, in addition to being

very complicated, they deviate considerably from the distributions of the non-circular statistics. This observation has already been made by Kendall (1954), who showed that whilst the expectation of the circular first lag serial correlation is given by

$$\mathscr{E}(\bar{r}_1)_{\text{c.}} = \alpha_1 - \frac{1}{n}(1+4\alpha_1) + O\left(\frac{1}{n^2}\right), \tag{7·5}$$

the corresponding formula for the non-circular coefficient is given by

$$\mathscr{E}(\bar{r}_1)_{\text{n.c.}} = \alpha_1 - \frac{1}{n}(1+3\alpha_1) + O\left(\frac{1}{n^2}\right). \tag{7·6}$$

It will be shown in this section that $\mathscr{E}_s(\bar{r}_1)_{\text{c.}}$ agrees to the required order with (7·6), and this illustrates the agreement to $O(n^{-1})$ between the smoothed circular and exact non-circular distributions.

(7·3) is the distribution first put forward by Leipnik (1947), and (7·4) is an analogous distribution with the serial correlation corrected for the mean—this agrees with the saddle-point approximation given by Daniels as equation (5·4).

The form for $p_s(\bar{r}_1 \mid \alpha_1 = 0)$ has been given in § 4, and it is to be observed that this function has a negative loop in the positive tail so that it is not a density function, although it integrates out to unity. For moderate n, however, it behaves like a type β distribution.

We now proceed to show that the moments of (7·4) may be expressed as functions of the moments of the Leipnik distribution. It is necessary first of all to derive the first four moments of the latter distribution. The first two moments were derived by Leipnik himself by a very complicated method starting from $p_s(r_1, c_0 \mid \alpha_1)$. A simpler method is required in order to derive the higher moments of $p_s(r_1 \mid \alpha_1)$ and, also, this must be capable of extension to cover the new case $p_s(\bar{r}_1 \mid \alpha_1)$.

This is obtained by expanding the distribution in the form of an infinite series in the following manner:

$$p_s(r_1 \mid \alpha_1) = p_s(r_1 \mid \alpha_1 = 0)(1+\alpha_1^2)^{-\frac{1}{2}n} \sum_{j=0}^{\infty} \frac{(\frac{1}{2}n+j-1)!}{(\frac{1}{2}n-1)!\,j!} \lambda^j r_1^j,$$

where $\lambda = 2\alpha_1/(1+\alpha_1^2)$, so that $|\lambda| < 1$, and hence

$$m'_{2k} = \mathscr{E}(r_1^{2k} \mid \alpha_1) = (1+\alpha_1^2)^{-\frac{1}{2}n} \sum_{j=0}^{\infty} \frac{(\frac{1}{2}n+2j-1)!}{(\frac{1}{2}n-1)!\,(2j)!} \lambda^{2j} \mu'_{2j+2k}, \tag{7·7}$$

$$m'_{2k-1} = \mathscr{E}(r_1^{2k-1} \mid \alpha_1) = (1+\alpha_1^2)^{-\frac{1}{2}n} \sum_{j=1}^{\infty} \frac{(\frac{1}{2}n+2j-2)!}{(\frac{1}{2}n-1)!\,(2j-1)!} \lambda^{2j-1} \mu'_{2j+2k-2}. \tag{7·8}$$

By substituting for the null moments, viz. those of the type α distribution, it is possible to derive the non-null moments by manipulation of the resulting infinite series and using the basic identity $m'_0 = 1$. The details are omitted here and the results quoted in the following form:

$$m'_1 = \frac{n\alpha_1}{n+2},$$

$$m'_2 = \frac{1}{n+2} + \frac{n(n+1)\alpha_1^2}{(n+2)(n+4)},$$

$$m'_3 = \frac{3n\alpha_1}{(n+2)(n+4)} + \frac{n(n+1)\alpha_1^3}{(n+4)(n+6)},$$

$$m'_4 = \frac{3}{(n+2)(n+4)} + \frac{6n(n+1)\alpha_1^2}{(n+2)(n+4)(n+6)} + \frac{n(n+1)(n+3)\alpha_1^4}{(n+4)(n+6)(n+8)}.$$

These lead to central moments in the form

$$m_2 = \frac{1}{n+2}\left\{1 - \frac{n(n-2)\,\alpha_1^2}{(n+2)\,(n+4)}\right\},$$

$$m_3 = \frac{-6n\alpha_1}{(n+2)^2\,(n+4)} + \frac{2n(n-2)\,(3n-2)\,\alpha_1^3}{(n+2)^3\,(n+4)\,(n+6)},$$

$$m_4 = \frac{3}{(n+2)\,(n+4)} - \frac{6n\alpha_1^4(n^2-8n-4)}{(n+2)^3\,(n+4)\,(n+6)} + \frac{3n\alpha_1^4(n-2)\,(n^3-14n^2+12n-8)}{(n+2)^4\,(n+4)\,(n+6)\,(n+8)}.$$

In a similar manner, (7·4) may be expanded in the form of an infinite series and the moments expressed in the form

$$\overline{m}'_i = \frac{(1+\alpha_1^2)^{-\frac{1}{2}n'}}{(1-\alpha_1)} \sum_{j=0}^{\infty} \frac{(\frac{1}{2}n'+j-1)!}{(\frac{1}{2}n'-1)!\,j!}\,\lambda^j \overline{\mu}'_{i+j}, \tag{7·9}$$

where $\overline{\mu}'_r$ are the moments of $p_s(\bar{r}_1 \mid \alpha_1 = 0)$ and are given by (4·1) and (4·2). Using the notation $m'_i(n')$ to denote the ith moment about zero of $p_s(r_1 \mid \alpha_1)$ but based on $n' = n-1$ instead of n observations, it follows that (5·1) and (5·2) may be written in the form

$$\left.\begin{aligned} \overline{\mu}'_{2k} &= \mu'_{2k}, \\ \overline{\mu}'_{2k-1} &= -\left(\frac{n-1}{n+1}\right)\mu'_{2k}. \end{aligned}\right\} \tag{7·10}$$

We now proceed to prove the following identity:

$$\overline{m}'_i = \frac{1}{(1-\alpha_1)}\left\{m'_i(n') - \left(\frac{n+1}{n-1}\right)m'_{i+1}(n')\right\}. \tag{7·11}$$

It is necessary to distinguish between the two cases where i is odd or even. Writing $i = 2k$ in (7·9), it follows that

$$\overline{m}'_{2k} = \frac{(1+\alpha_1^2)^{-\frac{1}{2}n'}}{(1-\alpha_1)}\left\{\sum_{j=0}^{\infty}\frac{(\frac{1}{2}n'-1+2j)!}{(\frac{1}{2}n'-1)!\,(2j)!}\,\lambda^{2j}\overline{\mu}'_{2j+2k} + \sum_{j=1}^{\infty}\frac{(\frac{1}{2}n'+2j-2)!}{(\frac{1}{2}n'-1)!\,(2j-1)!}\,\lambda^{2j-1}\overline{\mu}'_{2j+2k-1}\right\},$$

and substituting for the expressions given by (7·7) and (7·8) and using (7·9) and (7·10), we obtain (7·11). The proof for i odd is similar. Since the values of m'_i are known for $i = 1, 2, 3$ and 4, it follows that $\overline{m}'_i$ may be derived for $i = 1, 2, 3$. If the necessary substitution is made in (7·11), it may be shown that

$$\overline{m}'_1 = \alpha_1 - \frac{1}{n-1}(1+3\alpha_1) + \frac{4n\alpha_1(1-3\alpha_1)}{(n^2-1)\,(n+3)\,(1-\alpha_1)} + \frac{12\alpha_1}{(n^2-1)\,(n+3)}, \tag{7·12}$$

$$\overline{m}'_2 = \frac{1}{n+1} + \frac{n}{n+3}\alpha_1^2 - \frac{2n\alpha_1(1+2\alpha_1)}{(n^2-1)\,(n+3)} - \frac{16n}{(n+1)\,(n+3)\,(n+5)}\,\frac{\alpha_1^3}{(1-\alpha_1)}, \tag{7·13}$$

$$\begin{aligned}\overline{m}'_3 = &-\frac{3}{(n-1)\,(n+3)} + \frac{n}{n+3}\alpha_1^4 - \frac{2n}{(n+3)\,(n+5)}\alpha_1^2(1+4\alpha_1) \\ &- \frac{3n(n-3)}{(n^2-1)\,(n+3)}\,\frac{\alpha_1}{1-\alpha_1} + \frac{5\alpha_1^4}{1-\alpha_1}\,\frac{n(n+1)}{(n+3)\,(n+5)\,(n+7)}.\end{aligned} \tag{7·14}$$

These lead to the following asymptotic expansions for the mean and variance:

$$\overline{m}'_1 = \alpha_1 - \frac{1}{n}(1+3\alpha_1) - \frac{1}{n^2}\left\{(1-\alpha_1) + \frac{8\alpha_1^2}{1-\alpha_1}\right\} + O\left(\frac{1}{n^3}\right), \tag{7·15}$$

$$\overline{m}_2 = \frac{1}{n}(1-\alpha_1^2) + \frac{1}{n^2}\{-2+4\alpha_1+14\alpha_1^2\} + O\left(\frac{1}{n^3}\right). \tag{7·16}$$

The above results are interesting in that they show that

(i) $\overline{m}_1'$ agrees up to $O(n^{-1})$ with the first moment of the non-circular statistic independently given by Kendall (1954) and Marriot & Pope (1954).

(ii) The term in $O(n^{-2})$ makes a large contribution to the variance even for moderate α_1 (it may be seen that the contribution is positive if $\alpha_1 > 0{\cdot}26$). This is in agreement with the conclusions of Marriot & Pope (1954) who conducted sampling experiments with artificial A.R. (1) series and found that the expected variances to $O(n^{-1})$ were significantly less than the observed variances. The variance of the non-circular statistic is not known to $O(n^{-2})$ so that we suggest that formula (7·16) be used as an estimate.

Table 1

Source	α_1	n	k*	Observed variance	$\mathscr{E}$ (var.) to $O(n^{-1})$	$\mathscr{E}$ (var.) to $O(n^{-2})$	$P(\chi_1^2)$	$P(\chi_2^2)$
Marriot & Pope (1954)	0·4	20	60	0·0529	0·0420	0·0466	0·09	0·23
	0·4	40	40	0·0398	0·0210	0·0224	0·0007	0·002
	0·4	60	30	0·0258	0·0140	0·0145	0·004	0·007
	0·8	20	60	0·0419	0·0180	0·0434	<0·00001	0·55
	0·8	40	40	0·0156	0·0090	0·0154	0·003	0·43
	0·8	60	30	0·0141	0·0060	0·0088	<0·00001	0·02
Rao & Som (1951)	0·8	15	50	0·0509	0·0240	0·0696	<0·00001	0·91
	0·8	35	25	0·0212	0·0103	0·0186	0·002	0·28
Kendall (1949), Series 9 (extended)	0·7	22	40	0·0462	0·0232	0·0390	0·0002	0·20

* k = number of subsets on which the variances are based.

Table 1 shows the improvement in the estimates of variance when the term in $O(n^{-2})$ is taken into account. The observed variances of $\bar{r}_1$ in the first six series are those given by Marriot & Pope (1954, Tables 9 and 14). The investigation was supplemented by calculating the variance of $\bar{r}_1$ for the following series:

(*a*) Two Markoff series with $\alpha_1 = 0{\cdot}8$ given by Rao & Som (1951), the serial correlations being based on 50 sets of 15 items and 25 sets of 35 items respectively.

(*b*) 40 sets of 22 items from a series given by Kendall (1949) which was extended from 500 to 1000 terms by the author.

In all cases the serial correlations were non-circular but varied in definition slightly due to the use of a pooled or unpooled variance in the denominators. It may be seen that the agreement between observed and expected variances to $O(n^{-2})$ is good except in two cases.

The analysis of this section has been extended to the Yule scheme where the situation is more complex. Details of this work will be published at a later stage. The same effect has

been observed here as with the Markoff scheme, viz. that estimates of variances and covariances to $O(n^{-1})$ are inadequate for autocorrelated series.

In conclusion, I would like to record my thanks to the Department of Scientific and Industrial Research for a maintenance grant during the period of this research.

REFERENCES

ANDERSON, R. L. (1942). *Ann. Math. Statist.* **13**, 1.
DANIELS, H. E. (1956). *Biometrika*, **43**, 169.
DIXON, W. J. (1944). *Ann. Math. Statist.* **15**, 119.
JENKINS, G. M. (1954). *Biometrika*, **41**, 405.
KENDALL, M. G. (1948). *Advanced Theory of Statistics*, **1**. London: Charles Griffin and Co.
KENDALL, M. G. (1949). *Biometrika*, **36**, 267.
KENDALL, M. G. (1954). *Biometrika*, **41**, 403.
KOOPMANS, T. L. (1942). *Ann. Math. Statist.* **13**, 14.
LEIPNIK, R. L. (1947). *Ann. Math. Statist.* **18**, 80.
MADOW, W. G. (1945). *Ann. Math. Statist.* **16**, 308.
MARRIOT, F. H. C. & POPE, J. A. (1954). *Biometrika*, **41**, 390.
PITMAN, E. J. G. (1937). *Proc. Camb. Phil. Soc.* **33**, 212.
QUENOUILLE, M. H. (1947). *J. R. Statist. Soc.* **110**, 123.
QUENOUILLE, M. H. (1949). *Ann. Math. Statist.* **19**, 561.
RAO, S. R. & SOM, K. S. (1951). *Sankhyā*, **11**, 239.
RUDRA, A. (1952). *Biometrika*, **39**, 434.
RUDRA, A. (1954). *Calcutta Statist. Ass. Bull.* **5** (18), 59.
RUDRA, A. (1955). *Sankhyā*, **15**, 9.
WATSON, G. S. (1956). *Biometrika*, **43**, 161.
WHITTLE, P. (1952). *Biometrika*, **39**, 309.
WOLD, H. (1938). *A Study in the Analysis of Stationary Time Series*. Uppsala: Almquist and Wicksell.

MISCELLANEA

A class of distributions for which the maximum-likelihood estimator is unbiased and of minimum variance for all sample sizes

By D. E. BARTON

University College London

1. We consider a sample $(x_1, \ldots, x_n)$ randomly and independently drawn from a population whose law $p(x;\ \theta)$ depends on an unknown parameter θ in such a manner as to admit a sufficient statistic for θ.

Blackwell (1947) noted that if t' is an unbiased (but possibly inefficient) estimator of θ, then, if t is a statistic sufficient for θ,

$$T = \mathscr{E}(t' \mid t) \tag{1}$$

is purely a function of t (and therefore sufficient for θ) and

$$\mathscr{E}(T) = \theta. \tag{2}$$

Further, since

$$\operatorname{Var} t' = \operatorname{Var} T + \mathscr{E}\left[\operatorname{Var}(t' \mid t)\right] \tag{3}$$

he noted that T has a variance not greater than that of any unbiased estimator from which it may be derived. Hoel (1951) showed that a function of T, sufficient for θ and obeying (2), is effectively unique, and so we may talk of *the* best unbiased estimator.

Since (as is typified by the example Blackwell gives) it is generally easy to find *an* unbiased estimator of θ and the method of maximum likelihood yields a sufficient statistic, $\hat{\theta}$ say, for θ in these circumstances, it would seem that it is always possible to find the best unbiased estimator of θ. Unfortunately, $\mathscr{E}(t' \mid \hat{\theta})$ may be a function which it is not feasible to evaluate. For instance, in the example of equation (14) the sample arithmetic mean, $\bar{x}$, is unbiased whilst the sample geometric mean, g, is sufficient for θ, but it is not possible to evaluate the integral for $\mathscr{E}(\bar{x} \mid g)$. The present note remarks that it is always possible to choose a function $\phi = \phi(\theta)$ for which the maximum-likelihood estimate is the best unbiased estimator. The method is extended to the multi-parameter case.

2. Koopman (1936) and Pitman (1936) showed that the most general form of probability law admitting a sufficient statistic is

$$p(x;\ \theta) = \exp\{\alpha(\theta) f(x) + \beta(\theta) + g(x)\}, \tag{4}$$

where α, β, f, g are functions of the stated variables. The mean value of the log-likelihood function is thus

$$\overline{L} = \alpha(\theta)\overline{f} + \beta(\theta) + \overline{g}, \tag{5}$$

where

$$\overline{f} = \sum_{i=1}^{n} f(x_i)/n, \quad \overline{g} = \sum_{i=1}^{n} g(x_i)/n. \tag{6}$$

Hence

$$\frac{\partial \overline{L}}{\partial \theta} = \alpha'(\theta)\overline{f} + \beta'(\theta), \tag{7}$$

$$0 = \mathscr{E}\left\{\frac{\partial \overline{L}}{\partial \theta}\right\} = \alpha'(\theta)\,\mathscr{E}(\overline{f}) + \beta'(\theta), \tag{8}$$

and the maximum-likelihood equation is

$$0 = \left.\frac{\partial \overline{L}}{\partial \theta}\right|_{\theta=\hat{\theta}} = \alpha'(\hat{\theta})\overline{f} + \beta'(\hat{\theta}). \tag{9}$$

Thus if

$$\phi = \phi(\theta) = -\beta'(\theta)/\alpha'(\theta), \tag{10}$$

then

$$\mathscr{E}(\phi(\hat{\theta})) = \phi(\theta). \tag{11}$$

Expressing $p(x;\theta)$ as $p(x;\phi)$ by means of (11) we have for the maximum-likelihood estimator

$$\mathscr{E}(\hat{\phi}) = \phi.$$

Further, the form of $p(x;\phi)$ is such that the equality conditions for the Cramer–Rao inequality hold and so

$$\operatorname{Var}\hat{\phi} = -\left[n\mathscr{E}\left\{\frac{\partial^2}{\partial\phi^2}\log p(x;\phi)\right\}\right]^{-1} = [nA'(\phi)]^{-1}, \tag{12}$$

where $A(\phi)$ is the function α expressed as a function of ϕ by (11). Thus if

$$p(x;\theta) = \frac{x^\theta}{\Gamma(\theta+1)}e^{-x} \quad (x \geqslant 0) \tag{13}$$

$F(\theta)$ is the digamma function of θ and

$$\phi = \frac{d\log\Gamma(\theta+1)}{d\theta} = F(\theta), \tag{14}$$

then

$$\hat{\phi} = \frac{1}{n}\log(x_1 x_2 \ldots x_n), \tag{15}$$

and

$$\mathscr{E}(\hat{\phi}) = \phi, \quad \operatorname{Var}\phi = -F(\theta)/n. \tag{16}$$

It follows from the well-known theorem on sufficient estimators that not only the variance but also the characteristic function of $\hat{\phi}$ is known. Thus if $\psi(t)$ is the C.G.F. of $\hat{\phi}$,

$$\psi'(t) = i\phi\left[A + \frac{it}{n}\right] \tag{17}$$

whence, for instance, the shape constants for $\hat{\phi}$ are

$$\gamma_1 = -\frac{1}{n}\frac{A_3}{A_1^3}, \quad \gamma_2 = \frac{1}{n}\frac{3A_2^2 - A_1A_3}{A_1^2}, \tag{18}$$

where

$$A_r = \frac{d^rA(\phi)}{d\phi^r}. \tag{19}$$

3. Koopman and Pitman showed that the most general case of a law depending on p unknown parameters $(\theta_1, \ldots, \theta_p)$ admitting sufficient statistics had the form

$$p(x;\theta) = \exp\left\{\sum_{r=1}^{p}\alpha_r(\theta)f_r(x) + \beta(\theta) + g(x)\right\}, \tag{20}$$

where θ denotes $(\theta_1, \ldots, \theta_p)$. It follows as before that

$$0 = \sum_{r=1}^{p}\alpha'_{rs}(\theta)\,\mathscr{E}(\bar{f}_r) + \beta'_s(\theta) \quad (s = 1, \ldots, p) \tag{21}$$

and

$$0 = \sum_{r=1}^{p}\alpha'_{rs}(\hat{\theta})\bar{f}_r + \beta'_s(\hat{\theta}) \quad (s = 1, \ldots, p), \tag{22}$$

where

$$\alpha'_{rs}(\theta) = \frac{\partial}{\partial\theta_s}\alpha_r(\theta), \quad \beta'_s(\theta) = \frac{\partial}{\partial\theta_s}\beta(\theta), \tag{23}$$

and $\hat{\theta}$ denotes $(\hat{\theta}_1, \ldots, \hat{\theta}_p)$. Hence, if

$$\phi_r = \phi_r(\theta) = -B_r/A \quad (r = 1, \ldots, p), \tag{24}$$

where $A = |[\alpha'_{rs}]|$ and B_r is A with β'_s replacing α'_{rs} in the rth row, then (23) becomes

$$\bar{f}_r = \hat{\phi}_r \quad (r = 1, \ldots, p) \tag{25}$$

and (22) is

$$\mathscr{E}(\hat{\phi}_r) = \phi_r \quad (r = 1, \ldots, p). \tag{26}$$

The minimum variance property follows from Cramer's (1946) generalized inequality, namely, that, in his terminology, the concentration ellipsoid of $\hat{\phi}_1, \ldots, \hat{\phi}_p$ lies inside that of any other p estimators of $\phi_1, \ldots, \phi_p$.

The variance-covariance matrix of the $\{\hat{\phi}_r\}$ is easily seen to be

$$\frac{1}{n}\left[\frac{\partial\alpha_s}{\partial\phi_r}\right]^{-1} = \frac{1}{n}\left[\frac{\partial\phi_r}{\partial\alpha_s}\right], \tag{27}$$

where $\partial\alpha_s/\partial\phi_r$ denotes the partial differential of α_s (expressed as a function of $\phi_1, \ldots, \phi_p$) with respect to ϕ_r. The right-hand side of (21) gives a simpler form since it does not require matrix inversion.

This follows also from the generalization of (17); namely

$$\frac{\partial\psi}{\partial t_r} = i\phi_r^*\left(\alpha_j + \frac{it_j}{n}\right) \quad (r = 1, \ldots, p), \tag{28}$$

where ψ is the joint C.G.F. of $\{\hat{\phi}_r\}$ in terms of dummy variables $\{t_r\}$ and $\phi_r^*(\alpha_j)$ denotes the value of ϕ_r when the variables $\{\theta_j\}$ are written as functions of $\{\alpha_j\}$.

4. For example, the Pearson type III law of known origin

$$p(x;\, \theta_1, \theta_2) = \theta_1^{\theta_2+1}\frac{x^{\theta_2}e^{-x\theta_1}}{\Gamma(\theta_2+1)} \quad (0 \leqslant x),$$

if
$$\phi_1 = (\theta_2+1)/\theta_1, \quad \phi_2 = F(\theta_2) - \log\theta_1$$

yields explicit unbiased maximum-likelihood estimators of ϕ_1, ϕ_2 with variance-covariance matrix

$$\frac{1}{n}\begin{bmatrix} (\theta_2+1)/\theta_1^2 & 1/\theta_1 \\ 1/\theta_1 & F(\theta_2) \end{bmatrix}.$$

The type I law
$$p(x;\, \theta_1, \theta_2) = (1-x)^{\theta_1}x^{\theta_2}/B(\theta_1+1, \theta_2+1) \quad (0 \leqslant x \leqslant 1)$$

gives
$$\phi_1 = F(\theta_1) - F(\theta_2+\theta_1+1), \quad \phi_2 = F(\theta_2) - F(\theta_2+\theta_1+1),$$

with the variance-covariance matrix

$$\frac{1}{n}\begin{bmatrix} F(\theta_1) - F(\theta_2+\theta_1+1) & -F(\theta_2+\theta_1+1) \\ -F(\theta_2+\theta_1+1) & F(\theta_2) - F(\theta_2+\theta_1+1) \end{bmatrix}.$$

It should be noted that these results do not apply to the multi-parameter case where interest is focused on one of the $\{\theta_r\}$ say θ_1, and the variables are so integrated out that we are left with the distribution of one sufficient statistic dependent solely on θ_1. For instance, when the variables (x_i) are the bivariate observations (y_i, z_i) from the five-parameter normal surface and we derive the marginal law of r (the sample correlation coefficient) which is dependent solely on ρ (its population value), the method yields

$$\mathscr{E}\left\{\frac{r}{\sqrt{(1-r^2)}}\right\} = \frac{n-2}{n-3}\frac{\rho}{\sqrt{(1-\rho^2)}},$$

i.e. this relation results from the analogous equations to (8) and (9). It is essentially different to equation (11) though it has interesting similarity. This is a particular case of a general result to be published by B. I. Harley.

5. In conclusion, it may be said that this note gives a method for obtaining a best unbiased sufficient statistic in cases where Blackwell's method does not yield tractable results. It may only be reasonably applied when the property of unbiasedness is of more importance than the functional form of the parameter estimated.

REFERENCES

BLACKWELL, D. (1947). *Ann. Math. Statist.* **18**, 105.
CRAMER, H. (1946). *Skand. AktuarTidskr.* **29**, 85.
HOEL, P. G. (1951). *Ann. Math. Statist.* **22**, 299.
KOOPMAN, B. O. (1936). *Trans. Amer. Math. Soc.* **39**, 399.
PITMAN, E. J. G. (1936). *Proc. Camb. Phil. Soc.* **37**, 567.

Further critical values for the two-means problem

BY W. H. TRICKETT, B. L. WELCH AND G. S. JAMES

University of Leeds

Tables have already been given of *one-sided* 5 and 1 % critical values of a quantity, v, used in comparing two means, and more generally used in a wide class of comparisons whose accuracy depends in a certain way on two variances separately estimated (A. A. Aspin (1949), reproduced as Table 11 of Pearson & Hartley (1954)). We now present tables of *two-sided* 5 and 1 % critical values of v, i.e. the *one-sided* 2·5 and 0·5 % values.*

If y is a normally distributed estimate of a population parameter η with sampling variance $\lambda_1\sigma_1^2+\lambda_2\sigma_2^2$, where λ_1 and λ_2 are known positive constants, and if s_1^2 and s_2^2 are estimates of σ_1^2 and σ_2^2 distributed in the standard fashion with ν_1 and ν_2 degrees of freedom respectively, and if y, s_1^2 and s_2^2 are all independent, then we define

$$v = (y-\eta)/\sqrt{(\lambda_1 s_1^2+\lambda_2 s_2^2)}.$$

v has a distribution depending on the unknown ratio of σ_1^2 to σ_2^2, but, writing

$$c = \lambda_1 s_1^2/(\lambda_1 s_1^2+\lambda_2 s_2^2),$$

we may seek a function $V(c;\ \nu_1, \nu_2, \alpha)$, depending on the ratio of the observed s_1^2 and s_2^2, with the property that

$$\Pr\{v \geqslant V(c;\ \nu_1, \nu_2, \alpha)\} = \alpha,$$

where α is a prescribed probability. (Knowledge of such a function would permit us, for instance, to make confidence statements about η in the usual manner.) The formulation of the problem in these terms is due to Welch (1947).

In the present tables, and in the previous tables calculated by Mrs Aspin, the object has been to give values of $V(c;\ \nu_1, \nu_2, \alpha)$ to two decimal places. The methods of computation described by Welch do not allow this to be done for very small values of ν_1 and ν_2. In the present tables for $\alpha = 0{\cdot}025$ we have $\nu_1,\ \nu_2 \geqslant 8$, and for $\alpha = 0{\cdot}005$ we have $\nu_1, \nu_2 \geqslant 10$. Even with this limitation, in a few instances there is doubt about the second decimal place. We are, in fact, admitting the possibility of the occasional occurrence of errors up to 1 unit in the last figure, rather than the usual $\frac{1}{2}$ unit due to rounding off.

Where interpolation is necessary, *direct* linear interpolation should be used in all directions in the tables except in any panel which is bordered by ν_1 or $\nu_2 = \infty$. In such cases *harmonic* linear interpolation may be used.

For an example of the use of the tables reference should be made to Aspin (1949) or Pearson & Hartley (1954, p. 27).

Some of the calculations made in computing and checking the present tables were carried out on the Manchester University (MK.I) automatic electronic computer. We would like to acknowledge the assistance of Dr D. W. J. Cruickshank, Miss D. E. Pilling and Mr J. F. P. Donovan, of the Department of Inorganic and Structural Chemistry, Leeds University, in helping one of us (G.S.J.) to make use of this machine.

REFERENCES

ASPIN, A. A. (1949). Tables for use in comparisons whose accuracy involves two variances, separately estimated (with an Appendix by B. L. Welch). *Biometrika*, **36**, 290–6. Reprinted as Table 11 of Pearson & Hartley (1954).

PEARSON, E. S. & HARTLEY, H. O. (1954). *Biometrika Tables for Statisticians*, **1**. Cambridge: The University Press for the Biometrika Trustees.

WELCH, B. L. (1947). The generalization of 'Student's' problem when several different population variances are involved. *Biometrika*, **34**, 28–35.

* The new tables are arranged in precisely the same form as the earlier tables, but their descriptive title has been slightly modified.

*Upper $2\frac{1}{2}$% critical values of $v = (y-\eta)/\sqrt{(\lambda_1 s_1^2 + \lambda_2 s_2^2)}$ (i.e. upper 5% critical values of $|v|$)**

	$\frac{\lambda_1 s_1^2}{(\lambda_1 s_1^2 + \lambda_2 s_2^2)}$	0·0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	1·0
ν_2	ν_1											
8	8	2·31	2·25	2·20	2·14	2·10	2·08	2·10	2·14	2·20	2·25	2·31
	10	2·31	2·25	2·20	2·15	2·10	2·08	2·08	2·11	2·14	2·19	2·23
	12	2·31	2·25	2·20	2·15	2·10	2·07	2·07	2·08	2·11	2·14	2·18
	15	2·31	2·25	2·20	2·15	2·10	2·07	2·05	2·06	2·08	2·10	2·13
	20	2·31	2·25	2·20	2·15	2·10	2·06	2·04	2·04	2·05	2·07	2·09
	∞	2·31	2·25	2·20	2·14	2·09	2·05	2·01	1·99	1·97	1·96	1·96
10	8	2·23	2·19	2·14	2·11	2·08	2·08	2·10	2·15	2·20	2·25	2·31
	10	2·23	2·18	2·14	2·11	2·08	2·06	2·08	2·11	2·14	2·18	2·23
	12	2·23	2·18	2·14	2·10	2·07	2·06	2·06	2·08	2·11	2·14	2·18
	15	2·23	2·18	2·14	2·10	2·07	2·05	2·05	2·06	2·08	2·10	2·13
	20	2·23	2·18	2·14	2·10	2·07	2·05	2·04	2·04	2·05	2·06	2·09
	∞	2·23	2·18	2·14	2·10	2·06	2·03	2·00	1·98	1·97	1·96	1·96
12	8	2·18	2·14	2·11	2·08	2·07	2·07	2·10	2·15	2·20	2·25	2·31
	10	2·18	2·14	2·11	2·08	2·06	2·06	2·07	2·10	2·14	2·18	2·23
	12	2·18	2·14	2·11	2·08	2·06	2·05	2·06	2·08	2·11	2·14	2·18
	15	2·18	2·14	2·11	2·08	2·06	2·04	2·04	2·06	2·08	2·10	2·13
	20	2·18	2·14	2·11	2·08	2·05	2·04	2·03	2·03	2·05	2·06	2·09
	∞	2·18	2·14	2·11	2·07	2·04	2·02	1·99	1·98	1·97	1·96	1·96
15	8	2·13	2·10	2·08	2·06	2·05	2·07	2·10	2·15	2·20	2·25	2·31
	10	2·13	2·10	2·08	2·06	2·05	2·05	2·07	2·10	2·14	2·18	2·23
	12	2·13	2·10	2·08	2·06	2·04	2·04	2·06	2·08	2·11	2·14	2·18
	15	2·13	2·10	2·08	2·05	2·04	2·03	2·04	2·05	2·08	2·10	2·13
	20	2·13	2·10	2·08	2·05	2·04	2·03	2·03	2·03	2·05	2·06	2·09
	∞	2·13	2·10	2·07	2·05	2·02	2·00	1·99	1·97	1·97	1·96	1·96
20	8	2·09	2·07	2·05	2·04	2·04	2·06	2·10	2·15	2·20	2·25	2·31
	10	2·09	2·06	2·05	2·04	2·04	2·05	2·07	2·10	2·14	2·18	2·23
	12	2·09	2·06	2·05	2·03	2·03	2·04	2·05	2·08	2·11	2·14	2·18
	15	2·09	2·06	2·05	2·03	2·03	2·03	2·04	2·05	2·08	2·10	2·13
	20	2·09	2·06	2·05	2·03	2·02	2·02	2·02	2·03	2·05	2·06	2·09
	∞	2·09	2·06	2·04	2·02	2·01	1·99	1·98	1·97	1·96	1·96	1·96
∞	8	1·96	1·96	1·97	1·99	2·01	2·05	2·09	2·14	2·20	2·25	2·31
	10	1·96	1·96	1·97	1·98	2·00	2·03	2·06	2·10	2·14	2·18	2·23
	12	1·96	1·96	1·97	1·98	1·99	2·02	2·04	2·07	2·11	2·14	2·18
	15	1·96	1·96	1·97	1·97	1·99	2·00	2·02	2·05	2·07	2·10	2·13
	20	1·96	1·96	1·96	1·97	1·98	1·99	2·01	2·02	2·04	2·06	2·09
	∞	1·96	1·96	1·96	1·96	1·96	1·96	1·96	1·96	1·96	1·96	1·96

* y is normally distributed about η with variance $\lambda_1\sigma_1^2 + \lambda_2\sigma_2^2$ and s_1^2 and s_2^2 are independent estimates of σ_1^2 and σ_2^2, based on ν_1 and ν_2 degrees of freedom respectively. λ_1 and λ_2 are known constants.

In the problem of comparing the means of samples taken from two normal populations, put $y = (\bar{x}_1 - \bar{x}_2)$, $\nu_1 = (n_1 - 1)$, $\nu_2 = (n_2 - 1)$, $\lambda_1 = 1/n_1$ and $\lambda_2 = 1/n_2$, where n_1 and n_2 are the sample sizes.

*Upper $\frac{1}{2}$ % critical values of $v = (y-\eta)/\sqrt{(\lambda_1 s_1^2 + \lambda_2 s_2^2)}$ (i.e. upper 1 % critical values of $|v|$)**

	$\frac{\lambda_1 s_1^2}{(\lambda_1 s_1^2 + \lambda_2 s_2^2)}$	0·0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	1·0
ν_2	ν_1											
10	**10**	3·17	3·08	3·00	2·90	2·82	2·79	2·82	2·90	3·00	3·08	3·17
	12	3·17	3·08	3·00	2·91	2·82	2·78	2·79	2·84	2·91	2·98	3·05
	15	3·17	3·08	3·00	2·91	2·82	2·77	2·76	2·78	2·83	2·89	2·95
	20	3·17	3·08	3·00	2·91	2·82	2·76	2·73	2·74	2·76	2·80	2·85
	30	3·17	3·08	3·00	2·91	2·82	2·75	2·71	2·69	2·70	2·72	2·75
	∞	3·17	3·08	2·99	2·91	2·82	2·74	2·67	2·63	2·60	2·58	2·58
12	**10**	3·05	2·98	2·91	2·84	2·79	2·78	2·82	2·91	3·00	3·08	3·17
	12	3·05	2·98	2·91	2·84	2·78	2·76	2·78	2·84	2·91	2·98	3·05
	15	3·05	2·98	2·91	2·84	2·78	2·75	2·75	2·78	2·83	2·89	2·95
	20	3·05	2·98	2·91	2·84	2·78	2·74	2·72	2·73	2·76	2·80	2·85
	30	3·05	2·98	2·91	2·84	2·77	2·73	2·70	2·69	2·70	2·72	2·75
	∞	3·05	2·98	2·91	2·84	2·77	2·71	2·65	2·62	2·59	2·58	2·58
15	**10**	2·95	2·89	2·83	2·78	2·76	2·77	2·82	2·91	3·00	3·08	3·17
	12	2·95	2·89	2·83	2·78	2·75	2·75	2·78	2·84	2·91	2·98	3·05
	15	2·95	2·89	2·83	2·78	2·74	2·73	2·74	2·78	2·83	2·89	2·95
	20	2·95	2·89	2·83	2·78	2·74	2·71	2·71	2·73	2·76	2·80	2·85
	30	2·95	2·89	2·83	2·78	2·73	2·70	2·68	2·68	2·70	2·72	2·75
	∞	2·95	2·89	2·83	2·77	2·72	2·67	2·64	2·61	2·59	2·58	2·58
20	**10**	2·85	2·80	2·76	2·74	2·73	2·76	2·82	2·91	3·00	3·08	3·17
	12	2·85	2·80	2·76	2·73	2·72	2·74	2·78	2·84	2·91	2·98	3·05
	15	2·85	2·80	2·76	2·73	2·71	2·71	2·74	2·78	2·83	2·89	2·95
	20	2·85	2·80	2·76	2·73	2·70	2·70	2·70	2·73	2·76	2·80	2·85
	30	2·85	2·80	2·76	2·72	2·69	2·68	2·67	2·68	2·70	2·72	2·75
	∞	2·85	2·80	2·76	2·72	2·68	2·65	2·62	2·60	2·59	2·58	2·58
30	**10**	2·75	2·72	2·70	2·69	2·71	2·75	2·82	2·91	3·00	3·08	3·17
	12	2·75	2·72	2·70	2·69	2·70	2·73	2·77	2·84	2·91	2·98	3·05
	15	2·75	2·72	2·70	2·68	2·68	2·70	2·73	2·78	2·83	2·89	2·95
	20	2·75	2·72	2·70	2·68	2·67	2·68	2·69	2·72	2·76	2·80	2·85
	30	2·75	2·72	2·69	2·67	2·66	2·66	2·66	2·67	2·69	2·72	2·75
	∞	2·75	2·72	2·69	2·66	2·64	2·62	2·60	2·59	2·58	2·58	2·58
∞	**10**	2·58	2·58	2·60	2·63	2·67	2·74	2·82	2·91	2·99	3·08	3·17
	12	2·58	2·58	2·59	2·62	2·65	2·71	2·77	2·84	2·91	2·98	3·05
	15	2·58	2·58	2·59	2·61	2·64	2·67	2·72	2·77	2·83	2·89	2·95
	20	2·58	2·58	2·59	2·60	2·62	2·65	2·68	2·72	2·76	2·80	2·85
	30	2·58	2·58	2·58	2·59	2·60	2·62	2·64	2·66	2·69	2·72	2·75
	∞	2·58	2·58	2·58	2·58	2·58	2·58	2·58	2·58	2·58	2·58	2·58

* y is normally distributed about η with variance $\lambda_1\sigma_1^2 + \lambda_2\sigma_2^2$, and s_1^2 and s_2^2 are independent estimates of σ_1^2 and σ_2^2, based on ν_1 and ν_2 degrees of freedom respectively. λ_1 and λ_2 are known constants.

In the problem of comparing the means of samples taken from two normal populations, put $y = (\bar{x}_1 - \bar{x}_2)$, $\nu_1 = (n_1 - 1)$, $\nu_2 = (n_2 - 1)$, $\lambda_1 = 1/n_1$ and $\lambda_2 = 1/n_2$, where n_1 and n_2 are the sample sizes.

An approximation for the symmetric, quadrivariate normal integral

By J. A. McFADDEN

U.S. Naval Ordnance Laboratory, Silver Spring, Maryland

The generalized tetrachoric series for the multivariate normal integral, as given by A. C. Aitken (unpublished), Kendall (1941, 1945) and Moran (1948), is not considered useful for computation, as stated by David (1953). The purpose of this note is to sum this series approximately in one special case, namely, the case of four variables when all six correlation coefficients are equal.

Suppose that x_1, x_2, x_3 and x_4 obey a quadrivariate normal distribution and that all the off-diagonal elements of the correlation matrix are equal to ρ. Let $P_4(\rho)$ be the probability that all four variables are simultaneously positive. Following Moran's paper, we may write

$$P_4(\rho) = \frac{1}{16} + \frac{3}{4\pi}\sin^{-1}\rho + \frac{1}{4\pi^2}S(\rho), \tag{1}$$

where
$$S(\rho) = 3\rho^2 - 4\rho^3 + 7\rho^4 - 12\rho^5 + \tfrac{338}{15}\rho^6 - 44\rho^7 + \tfrac{3162}{35}\rho^8 - \ldots. \tag{2}$$

The series (2) appears hopeless for computation. Proceeding formally, we make the transformation

$$\rho = \sin\phi \quad (-\sin^{-1}\tfrac{1}{3} \leqslant \phi \leqslant \tfrac{1}{2}\pi), \tag{3}$$

and expand $S(\rho)$ in powers of ϕ; then

$$S(\rho) = 3\phi^2 - 4\phi^3 + 6\phi^4 - 10\phi^5 + 18\phi^6 - \tfrac{1933}{30}\phi^7 + \tfrac{346}{5}\phi^8 - \ldots. \tag{4}$$

The series (4) can be summed approximately with the aid of a non-linear sequence-to-sequence transformation given by Shanks (1955). If A_0, A_1, A_2, A_3 and A_4 are the first five elements of a sequence, then Shanks's second-order transform e_2 provides a new sequence, the first element of which is

$$e_2(A_2) = \begin{vmatrix} A_0 & A_1 & A_2 \\ \Delta A_0 & \Delta A_1 & \Delta A_2 \\ \Delta A_1 & \Delta A_2 & \Delta A_3 \end{vmatrix} \Bigg/ \begin{vmatrix} 1 & 1 & 1 \\ \Delta A_0 & \Delta A_1 & \Delta A_2 \\ \Delta A_1 & \Delta A_2 & \Delta A_3 \end{vmatrix}, \tag{5}$$

where $\Delta A_n = A_{n+1} - A_n$.

Table 1

$1/\rho$	$P_4(\rho)$ from (6)	Ruben's $\overline{u}_4(1/\rho)$
1	0·50121 388	0·50000 00000
2	0·20000 802	0·20000 00000
3	0·14973 847	0·14973 76529
4	0·12647 941	0·12647 9249
5	0·11301 258	0·11301 25446
6	0·10422 406	0·10422 4047
7	0·09803 672	0·09803 672
8	0·09344 504	0·09344 505
9	0·08990 265	0·08990 27
10	0·08708 697	0·08708 71
11	0·08479 535	0·08479 5
12	0·08289 404	0·08289 4

Again proceeding formally, we apply e_2 to the first five partial sums of the series (4) and replace $S(\rho)$ by the first element of the transformed sequence. Then the result for the probability (1) is

$$P_4(\rho) \cong \frac{1}{16} + \frac{3}{4\pi}\phi + \frac{1}{4\pi^2}\frac{\phi^2(3+5\phi)}{(1+\phi)(1+2\phi)}, \tag{6}$$

where ϕ is defined by equation (3).

Since we have not justified the previous steps mathematically, we must test the result numerically. Table 1 gives a comparison of the approximation (6) with the results of the numerical integration of Ruben (1954) for the cases $1/\rho = 1, 2, \ldots, 12$. [If $\rho = 0$, equation (6) is exact.]

The agreement in Table 1 is exceptionally good except when ρ is near unity. If greater accuracy is desired, it may be obtained by the use of *seven* terms of equation (4) and the application of more complicated non-linear transformations, as given by Shanks (1955).*

The approximation (6) is considerably more accurate than another one described elsewhere by the author (McFadden, 1955), which was based on an analogy with Pólya's urn scheme; yet the formula (6) is no more complicated than the other.

If the multivariate normal integral is known for four variables, the result for five follows immediately, as shown by David (1953).

Attempts to extend the present method to n variables, with all correlation coefficients equal, and also to the *general* quadrivariate case (with correlation coefficients unequal) have been unsuccessful.

REFERENCES

David, F. N. (1953). *Biometrika*, **40**, 458.
Kendall, M. G. (1941). *Biometrika*, **32**, 196.
Kendall, M. G. (1945). *J. R. Statist. Soc.* **108**, 93.
McFadden, J. A. (1955). *Ann. Math. Statist.* **26**, 478.
Moran, P. A. P. (1948). *Biometrika*, **35**, 203.
Ruben, H. (1954). *Biometrika*, **41**, 200.
Shanks, D. (1955). *J. Math. Phys.* **34**, 1. (This work was first reported in 1949 in a Naval Ordnance Laboratory Memorandum.)

Weighted probits allowing for a non-zero response in the controls

By M. J. R. HEALY

Rothamsted Experimental Station

Black (1950) has shown how standard probit calculations can be carried out using tables of 'weighted probits'. These functions are defined by

$$\pi_0, \quad \text{weighted minimum probit} \equiv w(Y - P/Z),$$

$$\pi_1, \quad \text{weighted maximum probit} \equiv w(Y + Q/Z),$$

where Y is the probit value corresponding to a proportion P, Z is the corresponding ordinate, $w = Z^2/(PQ)$ is the probit weighting function and $Q = 1 - P$. If r subjects are observed to respond and s not to respond, the appropriate weight is $rw + sw$ and the working probit multiplied by the weight is $r\pi_1 + s\pi_0$. Since r and s are usually fairly small integers these are simple computations, and the necessary table look-ups are more convenient than those required when using the tables of Finney (1952) or Fisher & Yates (1953).

Weighted probits have certain advantages when probit calculations are carried out on high-speed automatic computers. In desk computation, the working probit is usually found by adding a fraction of the range $(1/Z)$ to the minimum working probit $(Y - P/Z)$. Neither of these functions is particularly well behaved, each becoming numerically large for extreme values of Y, and Finney (1952) has provided an extensive table of double entry to by-pass this stage of the computation. This table is too large to be held in the store of present-day machines. By contrast, the weighted probit functions behave reasonably from a numerical point of view in the range of practical importance, and, since high-order interpolation is a fast and simple process on automatic computers, wide-interval tables with comparatively few entries can be used which can be accommodated in the store of the machine without difficulty. It is of course possible to avoid the use of tables altogether by computing the functions *ab initio* as required, but in the present instance the necessary routines would be rather complicated and would probably occupy almost as much room in the store as the tables. The process would also be more time-consuming than a simple table look-up. A further advantage in the use of tables is that the graduating function

* The author (J.A.M.) recommends the use of the iterated first-order transformation for each numerical value of ϕ.

can be changed from probits to, for example, logits or angles without any alteration in the main programme, simply by inserting the appropriate tables.

An extension of the simple probit technique is required when it is necessary to allow for a non-zero response rate untreated material. Finney (1952, pp. 88–91) has shown that when this 'natural response rate' is well determined and not too large, its effect can be allowed for by two simple modifications to the ordinary process. In the first place, the observed proportions responding, P, are adjusted by the so-called Abbot's correction to give adjusted proportions $P' = (P-C)/(1-C)$, where C is the natural response rate; secondly, the weights have to be multiplied by a factor $P'/\{P'+C/(1-C)\}$. It is of some interest that these alterations can be taken care of by modifying the tables of weights and weighted probits, thus enabling the original machine programme to be used unchanged.

It is somewhat simpler to work in terms of Normal Equivalent Deviates rather than probits, the addition of 5 being in any case irrelevant when working on automatic computers. The necessary alterations to the weights, w, are taken care of by inserting a suitably modified table. The working probit, y, is given by

$$y = Q'y_0 + P'y_1,$$

where P' is the adjusted proportion responding, $Q' = 1-P'$ and y_0, y_1 are the minimum and maximum working probits. In terms of the observed proportion responding, this gives

$$y = \frac{Q}{1-C}y_0 + \frac{P-C}{1-C}y_1,$$

so that the required value, nwy, is given in terms of the weighted probits by

$$nwy = s\frac{\pi_0}{1-C} + r\frac{\pi_1}{1-C} - \frac{(r+s)C}{1-C}\pi_1$$

$$= s\left(\frac{\pi_0 - C\pi_1}{1-C}\right) + r\pi_1.$$

Thus the form of the computation remains unaltered if we use the adjusted weighted probits

$$\pi'_0 = (\pi_0 - C\pi_1)/(1-C)$$

and π_1, where the modified weights are used in computing π_0 and π_1 as functions of Y.

Tables of w, $\frac{1}{8}\pi'_0$ and $\frac{1}{8}\pi_1$ have been prepared covering the range of N.E.D.'s $-4(0{\cdot}125)+4$, with $C = 0(1)\,9\,\%$. Four-point interpolation in these tables (a convenient method is that described in Fisher & Yates (1953, p. 33)) gives 5-decimal accuracy over almost the whole range. Similar tables for the logit and angle function are in preparation.

REFERENCES

Black, A. N. (1950). Weighted probits and their use. *Biometrika*, **37**, 158–67.

Finney, D. J. (1952). *Probit Analysis*, 2nd ed. Cambridge University Press.

Fisher, R. A. & Yates, F. (1953). *Statistical Tables for Biological, Agricultural and Medical Research*, 4th ed. Edinburgh: Oliver and Boyd.

Treatment variances for experimental designs with serially correlated observations

By J. C. Butcher

Applied Mathematics Laboratory, New Zealand Department of Scientific and Industrial Research

1. Introduction and summary

Williams (1952) has considered the design of field experiments in which the fertilities of neighbouring plots are assumed to follow a linear autoregressive scheme of order one or two. Here a formal generalization is made and a notation introduced which simplifies the calculation of variances of estimates.

A model of the following form is assumed:

$$x_t + a_1 x_{t-1} + a_2 x_{t-2} + \ldots + a_p x_{t-p} = \epsilon_t, \tag{1}$$

where the ϵ_t are independent and distributed $N(0, \sigma^2)$. For convenience this will be written as

$$\sum_{i=0}^{p} a_i x_{t-i} = \epsilon_t, \tag{2}$$

it being understood that a_0 is to be replaced by 1 when written in full.

The yield is assumed to be equal to the sum of x_t and a term depending on the treatment, i.e.

$$y_t = x_t + b_{(t)}, \tag{3}$$

where the notation (t) denotes the number of the treatment used on the tth plot. There are c treatments.

Estimation equations are derived, and it is found that the information matrix factorizes in a manner (20) which simplifies the calculation of the variance of the difference of the estimates of treatment constants. Some examples are given.

2. Notation

A square matrix of order c all of whose elements are unity will be denoted by E. Consider matrices of the form $\alpha I + \beta E$. As $E^2 = cE$ we see that

$$(\alpha_1 I + \beta_1 E)(\alpha_2 I + \beta_2 E) = \alpha_1 \alpha_2 I + (\alpha_1 \beta_2 + \alpha_2 \beta_1 + c\beta_1 \beta_2) E. \tag{4}$$

If $\alpha_2 I + \beta_2 E$ is the inverse of $\alpha_1 I + \beta_1 E$, then

$$\alpha_2 = \frac{1}{\alpha_1}, \quad \beta_2 = \frac{-\beta_1}{\alpha_1(\alpha_1 + c\beta_1)}. \tag{5}$$

We are interested in the cases where the information matrix is $\alpha_1 I + \beta_1 E$ and the variance-covariance matrix $\alpha_2 I + \beta_2 E$. The variances will equal $\alpha_2 + \beta_2$ and the covariances will all equal β_2. Thus the variances of differences will equal

$$2\alpha_2 = 2/\alpha_1. \tag{6}$$

The symbol δ_i^j is defined for the ith plot and the jth treatment as unity if that plot receives this treatment and zero otherwise. As each plot receives some treatment we have

$$\sum_j \delta_i^j = 1. \tag{7}$$

Also we have such results as

$$\sum_{(t-j)=i} a_j = \sum_j a_j \delta_{t-j}^j. \tag{8}$$

The number of plots receiving the jth treatment is

$$\sum_i \delta_i^j. \tag{9}$$

3. Estimation of parameters

The frequency function of the ϵ's is

$$\prod_t \left[\frac{1}{\sqrt{(2\pi\sigma^2)}} e^{-\frac{1}{2}\epsilon_t^2/\sigma^2} \right]. \tag{10}$$

Since the x's and ϵ's are connected by the linear relation (1), the likelihood of the sample $(x_1, x_2, \ldots, x_n)$ can be directly calculated from (10). The Jacobian of the transformation is unity. In order to avoid end effects which introduce irrelevant complications, we shall regard the values of $x_0, x_{-1}, \ldots, x_{-p+1}$ as additional parameters to be estimated. Estimation of these by maximum likelihood immediately gives $\epsilon_1 = \epsilon_2 = \ldots = \epsilon_p = 0$. To see the effect of correlations between the estimates of these x's and of the other parameters, consider instead the $\epsilon_t (t = 1, 2, \ldots, p)$ which constitute independent linear combinations of the $x_t (t = 0, -1, \ldots, -p+1)$. We see that the estimates of these ϵ_t are independent of the estimates of the other parameters.

We may thus maximize the expression for the conditional likelihood

$$-L = \frac{1}{2} \sum_{p+1}^{n} \epsilon_t^2/\sigma^2 + n \log \sigma \tag{11}$$

in order to estimate the parameters a_i, b_j. σ^2 is, of course, best estimated by dividing the residual sum of squares by $n - 2p - c$.

The estimates of the remaining parameters θ are the solutions of

$$-\frac{\partial L}{\partial \theta} = \frac{1}{\sigma^2} \sum_{p+1}^{n} \epsilon_t \frac{\partial \epsilon_t}{\partial \theta} = 0. \tag{12}$$

This gives for $\theta = a_i$

$$\sum_{p+1}^{n} (y_{t-i} - b_{(t-i)}) \epsilon_t = 0, \tag{13}$$

and for $\theta = b_j$

$$\sum_{p+1}^{n} \left(\sum_{(t-i)=j} a_i \right) \epsilon_t = 0. \tag{14}$$

By (8) this can be written

$$\sum_{t=p+1}^{n} \sum_{i=0}^{p} (a_i \delta^j_{t-i} \epsilon_t) = 0. \tag{15}$$

These equations can be solved by successive approximation, taking the first approximation for the b_j as the mean yield for the plots receiving the jth treatment. Using these values the a_i can be approximated and their values used to improve the approximation of the b_j, and so forth.

4. Variances and covariances of estimates of treatment constants

The variance-covariance matrix of parameter estimates is the inverse of the information matrix, which s approximately given by

$$\left[E\left(-\frac{\partial^2 L}{\partial \theta_i \partial \theta_j} \right) \right]. \tag{16}$$

We observe that if $\theta_1 = b_j$, $\theta_2 = a_i$ or σ^2, then

$$E\left(-\frac{\partial^2 L}{\partial \theta_1 \partial \theta_2} \right) = 0. \tag{17}$$

We may thus consider merely the matrix

$$\left[E\left(-\frac{\partial^2 L}{\partial b_i \partial b_j} \right) \right] \tag{18}$$

as the information matrix for the b's.

Let V be the variance-covariance matrix for the b's. We have

$$V^{-1} = \left[\frac{1}{\sigma^2} \sum_{t=p+1}^{n} \{ \left(\sum_{(t-k)=i} a_k \right) \left(\sum_{(t-l)=j} a_l \right) \} \right]. \tag{19}$$

It is easily verified that

$$V^{-1} = (\Delta A)(\Delta A)'/\sigma^2, \tag{20}$$

where

$$\Delta = (\delta^i_j) \qquad (i = 1, 2, \ldots, c;\ j = 1, 2, \ldots, n), \tag{21}$$

$$A = (a_{x-i+j}) \quad (i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, n-p), \tag{22}$$

where a_k is set equal to zero if $k < 0$ or $> p$. That is, the information matrix can be expressed in terms of Δ which depends only on the design (and which characterizes the design) and on A, which depends only on the autoregressive coefficients. It seems reasonable to require that the design be symmetric with respect to all treatments, which leads to the requirement that V^{-1} be of the form $\alpha I + \beta E$. One must thus place certain restrictions on Δ. It is easily seen that the value of $n-p$ must equal mc (m being an integer, the number of times each treatment occurs). As the last p treatments must be the same as the first p, we need only consider the mc treatments. In all that follows it is assumed that an additional p plots are used. It is clear that a repetition of a suitable design of mc plots any number of times to make a design of length Nmc is also suitable.

Conditions for suitable designs are easily seen. Let $u^{(r)}_{ij}$ be the number of times the ith and the jth treatment occur r plots apart. Then the condition is that

$$(u^{(r)}_{ij}) = \alpha_r I + \beta_r E \tag{23}$$

for all r from 1 to p.

Suitable designs are given by Williams (1952) for the cases:

$$p = 1, \quad \alpha_1 = 0; \qquad p = 1, \quad \alpha_1 + \beta_1 = 0; \qquad p = 2, \quad \alpha_1 + \beta_1 = \alpha_2 + \beta_2 = 0,$$

which he calls respectively types (II b), (II a) and (III).

It is clear that when designs suitable for $p = 1$ are found we can generalize them to any p by repeating each treatment over p adjacent plots, e.g.

$$1212\ldots \quad \text{would become}$$
$$1111\ldots 1,\ 222\ldots 2,\ 111\ldots 1,\ 222\ldots 2, \ldots.$$

As designs for first-order autoregression where $\alpha_1 + \beta_1 = 0$ are easily constructed by writing a suitable succession of permutations of the c treatments, we thus have a method of finding designs for any number of lags.

As an example of the calculation of these matrices consider the case $c = 3$, $p = 2$, $n = 3m + 2$:

$$\Delta = \begin{bmatrix} 1 & . & . & 1 & . & . & \ldots \\ . & 1 & . & . & 1 & . & \ldots \\ . & . & 1 & . & . & 1 & \ldots \end{bmatrix}.$$

ΔA becomes

$$\begin{bmatrix} a_2 & 1 & a_1 & a_2 & \ldots \\ a_1 & a_2 & 1 & a_1 & \ldots \\ 1 & a_1 & a_2 & 1 & \ldots \end{bmatrix}.$$

Thus
$$V^{-1} = \frac{m}{\sigma^2}(1 + a_1^2 + a_2^2)\, I + \frac{m}{\sigma^2}(a_1 + a_2 + a_1 a_2)\,(E - I)$$
$$= \frac{m}{\sigma^2}(1 + a_1^2 + a_2^2 - a_1 - a_2 - a_1 a_2)\, I + \frac{m}{\sigma^2}(a_1 + a_2 + a_1 a_2)\, E.$$

Therefore the variance of the estimates of treatment differences, denoted by v, equals

$$v = \frac{2\sigma^2}{m} \frac{1}{1 - a_1 - a_2 + a_1^2 - a_1 a_2 + a_2^2} \tag{24}$$

by (6).

In the following examples the number of plots is $N + p$. $\rho = -a_1$ is written when $p = 1$.

Case: $p = 1$, $c = 2$:

Design	$v = \operatorname{var}(\hat{b}_1 - \hat{b}_2)$	
[121212...]	$\dfrac{4\sigma^2}{N(1+\rho)^2}$,	(25)
[11221122...]	$\dfrac{4\sigma^2}{N(1+\rho^2)}$.	(26)

The relative efficiency of (25) and (26) is $1 + 2\rho/(1 + \rho^2)$ which is greater than one if ρ is positive.

Case: $p = 2$, $c = 2$:

Design	$v = \operatorname{var}(\hat{b}_1 - \hat{b}_2)$	
[121212...]	$\dfrac{4\sigma^2}{N(1 - a_1 + a_2)^2}$,	(27)
[11221122...]	$\dfrac{4\sigma^2}{N(1 + a_1^2 - 2a_1 a_2 + a_2^2)}$.	(28)

Consider now a more general case:

$$\underbrace{111\ldots1}_{q_1}\ \underbrace{222\ldots2}_{r_1}\ \underbrace{111\ldots1}_{q_2}\ \underbrace{222\ldots2}_{r_2}\ \ldots\ \underbrace{111\ldots1}_{q_s}\ \underbrace{222\ldots2}_{r_s},$$

where $p \leqslant q_1, r_1 \ldots$; $\Sigma q = \Sigma r = s/\alpha = \frac{1}{2}N$. We have

$$v = \frac{4\sigma^2}{N\left[\left(\sum_{i=0}^{p} a_i\right)^2 - 2\alpha \sum_{i=0}^{p} \sum_{j=0}^{p} a_i a_j\, |\, i - j\, |\right]}. \tag{29}$$

Thus if $\Sigma\Sigma a_i a_j\, |\, i - j\, | > 0$ it is advantageous to make the runs of 1's and 2's as long as possible; otherwise the q's and r's should all equal p.

Case: $c = 3$, $p = 2$, $n = N+2$ (for case $p = 1$, $n = N+1$ substitute $a_2 = 0$, $a_1 = -\rho$):

Design	$v = \text{var}(\hat{b}_i - \hat{b}_j)$	
[123, 123 ...]	$\dfrac{6\sigma^2}{N[1-a_1-a_2+a_1^2-a_1a_2+a_2^2]}$,	(30)
[123, 312, 231 ...]	$\dfrac{6\sigma^2}{N[1-a_2+a_1^2+a_2^2]}$,	(31)
[123, 231, 312 ...]	$\dfrac{6\sigma^2}{N[1-a_1+a_2+a_1^2-a_1a_2+a_2^2]}$,	(32)
[11, 22, 33 ...]	$\dfrac{12\sigma^2}{N[2+a_1+2a_1^2+a_1a_2-2a_2+2a_2^2]}$.	(33)

The last is a special case of the following:

$$\underbrace{111\ldots1}_{p}\quad\underbrace{222\ldots2}_{p}\quad\underbrace{333\ldots3}_{p}\ldots$$

$$v = \frac{6p\sigma^2}{N\left[p\left(\sum_0^p a_i\right)^2 - \frac{3}{2}\sum_{i=0}^{p}\sum_{j=0}^{p} a_i a_j\,|\,i-j\,|\right]}. \tag{34}$$

Cases (24), (25), (26) and (30) are included in formulae given by Williams.

In conclusion I wish to express my thanks to Dr P. Whittle who aroused my interest in this subject.

REFERENCE

Williams, R. M. (1952). Experimental designs for serially correlated observations. *Biometrika*, **39**, 151.

The multivariate distribution of complex normal variables

By R. A. Wooding

Department of Scientific and Industrial Research, Wellington, New Zealand

Introduction

The expression for the distribution of several normal real variables with correlation was introduced for the first time by Bravais (1846). Since then it has been discussed by many writers, notably Pearson (1896) and Cramér (1937, Chapter X).

Its applications include the case when the components of the multivariate distribution are real stochastic time variables which obey the normal law, as in random noise (Rice, 1944, 1945).

However, for certain problems connected with the envelope of a random noise signal, it is advantageous to consider a complex form, of which the signal is a component (Bunimovitch, 1949). A logical step is to derive the equivalent distribution of an ensemble of complex variates for which the real and imaginary components are normally distributed. It is found that, provided the real and imaginary components x_n, y_n obey the covariance relations

$$E(x_m x_n) = E(y_m y_n), \quad E(x_m y_n) = -E(x_n y_m),$$

the distribution of N complex variates v_n has the simple form

$$\pi^{-N}\,|\,L\,|^{-1}\exp(-V'^{*}L^{-1}V),$$

where L is the Hermitian covariance matrix.

Since this does not appear to have been published before, a derivation based on the real-variable expression is given here.

Definition of complex variates

Consider N pairs of real variables $x_n(t)$, $y_n(t)$, $(1 \leqslant n \leqslant N)$, where t represents a sampling parameter. Let the complex variable

$$v_n(t) = x_n(t) + iy_n(t) \tag{1}$$

be such that it can be written in the form

$$v_n(t) = \sum_j (a_{nj} - ib_{nj}) \exp\{i\theta_j(t)\}, \tag{2}$$

in which the a_{nj}, b_{nj} are real coefficients. This complex Fourier series arises in numerous fields, particularly in theory related to time series.

From (1) and (2) we see that

$$x_n(t) = \sum_j \{a_{nj} \cos\theta_j(t) + b_{nj} \sin\theta_j(t)\},$$

$$y_n(t) = \sum_j \{-a_{nj} \sin\theta_j(t) + b_{nj} \cos\theta_j(t)\};$$

i.e. the x_n, y_n are in 'phase quadrature'. This property is expressed compactly by the reciprocal Hilbert transformations (Bunimovitch, 1949, p. 1231)

$$y_n(t) = \frac{1}{\pi} \mathop{\mathrm{C}\!\!\!\!\int}_{-\infty}^{\infty} \frac{d\sigma}{\sigma} x_n(t-\sigma), \tag{3a}$$

$$x_n(t) = \frac{1}{\pi} \mathop{\mathrm{C}\!\!\!\!\int}_{-\infty}^{\infty} \frac{d\sigma}{\sigma} y_n(t+\sigma), \tag{3b}$$

where $\mathop{\mathrm{C}\!\!\!\!\int}$ signifies the Cauchy principal value of the integral.

Covariance Relations

Using (3a, 3b), it is readily shown that the x_n, y_n satisfy the following covariance relations, which are fundamental:

$$\begin{aligned} E(y_m y_n) &= \frac{1}{\pi} \mathop{\mathrm{C}\!\!\!\!\int}_{-\infty}^{\infty} \frac{d\sigma}{\sigma} E[x_m(t-\sigma)\, y_n(t)] \\ &= \frac{1}{\pi} \mathop{\mathrm{C}\!\!\!\!\int}_{-\infty}^{\infty} \frac{d\sigma}{\sigma} E[x_m(t)\, y_n(t+\sigma)] \\ &= E(x_m x_n). \end{aligned} \tag{4a}$$

Similarly,

$$E(x_m y_n) = -E(x_n y_m). \tag{4b}$$

It will be seen that, if the relations (4a, 4b), together with the standardizing relations $E(x_n) = E(y_n) = 0$, are taken as the initial postulate, our definition of the complex variates will not involve Fourier series concepts.

If now we consider the complex column vector

$$V = X + iY \tag{5}$$

formed from the v_n in (1), the composite column vector $\{XY\}$ will have a covariance matrix of form

$$\Lambda = \begin{pmatrix} A & B \\ B' & A \end{pmatrix}, \tag{6}$$

where the submatrix A is symmetric and B is skew-symmetric. It is easily verified that a decomposition of the corresponding inverse matrix is

$$\begin{pmatrix} A & B \\ B' & A \end{pmatrix}^{-1} = \begin{pmatrix} (A + BA^{-1}B)^{-1} & \cdot \\ \cdot & (A + BA^{-1}B)^{-1} \end{pmatrix} \begin{pmatrix} A & -B \\ -B' & A \end{pmatrix} \begin{pmatrix} A^{-1} & \cdot \\ \cdot & A^{-1} \end{pmatrix}, \tag{7}$$

which is of the same form as (6), so we can write

$$\Lambda^{-1} = \begin{pmatrix} A & B \\ B' & A \end{pmatrix}^{-1} = \begin{pmatrix} P & Q \\ Q' & P \end{pmatrix}, \tag{8}$$

with
$$P = (A + BA^{-1}B)^{-1}, \tag{9a}$$

$$Q = -(A + BA^{-1}B)^{-1}BA^{-1}, \tag{9b}$$

and
$$P = P', \quad Q = -Q'. \tag{9c}$$

Then it follows from formula (9) that
$$P - iQ = (A - iB)^{-1}. \tag{10}$$

Normal Variates

When the real vectors X, Y are normal, the likelihood function of the composite vector $\{XY\}$ is

$$(2\pi)^{-N} |\Lambda|^{-\frac{1}{2}} \exp[-\tfrac{1}{2}(X'Y')\Lambda^{-1}\{XY\}], \tag{11}$$

for which the exponent may be written

$$-\tfrac{1}{2}(X'Y')\Lambda^{-1}\{XY\} = -\tfrac{1}{2}(X' - iY')(P - iQ)(X + iY), \tag{12}$$

since Q is skew-symmetric, while the decomposition

$$\begin{pmatrix} P & Q \\ Q' & P \end{pmatrix} = \begin{pmatrix} I & -iI \\ . & I \end{pmatrix}\begin{pmatrix} P - iQ & . \\ -Q & I \end{pmatrix}\begin{pmatrix} I & iI \\ . & P + iQ \end{pmatrix}$$

shows that
$$|\Lambda|^{-1} = |P - iQ|\,|P + iQ| = |P - iQ|^2 = |A - iB|^{-2}. \tag{13}$$

Combining the expressions (10)–(13), we now obtain the result

$$(2\pi)^{-N} |A - iB|^{-1} \exp[-\tfrac{1}{2}(X' - iY')(A - iB)^{-1}(X + iY)] \tag{14}$$

for the likelihood of the complex normal vector $X + iY$. But since the covariance matrix of the complex vector $V = X + iY$ is the Hermitian matrix

$$L = E(VV'^*) = 2(A - iB), \tag{15}$$

it follows that
$$|L| = 2^N |A - iB|, \tag{16}$$

and that (14) can be expressed in the alternative form

$$\pi^{-N} |L|^{-1} \exp(-V'^* L^{-1} V) \tag{17}$$

for the normal probability function of the N complex numbers v_n.

It will be noted that this simple result is true only if the relations (4a) and (4b) hold.

The Characteristic Function

For the density function (11), the characteristic function can be written

$$(2\pi)^{-N} |\Lambda|^{-\frac{1}{2}} \int_X \int_Y \exp[i(R'S')\{XY\} - \tfrac{1}{2}(X'Y')\Lambda^{-1}\{XY\}]\,dX\,dY, \tag{18}$$

if $dX\,dY = \prod_{n=1}^{N} dx_n\,dy_n$ and $\{RS\}$ is a real column vector. This becomes

$$\exp[-\tfrac{1}{2}(R'S')\Lambda\{RS\}] = \exp[-\tfrac{1}{2}(R' - iS')(A - iB)(R + iS)], \tag{19}$$

since B is skew-symmetric. Hence, in the complex-variate notation (18) and (19) give

$$\pi^{-N} |L|^{-1} \int_V \exp[i\,\mathrm{Rl}(T'^* V) - V'^* L^{-1} V]\,dV = \exp(-\tfrac{1}{4} T'^* L T), \tag{20}$$

where $dV = dX\,dY$, $T = R + iS$, and Rl signifies that the real part is taken.

It is evident from the above that the theory of the distribution of complex variates is most readily deduced from real variate theory.

The author wishes to acknowledge the valuable advice and suggestions of Dr P. Whittle, of Applied Mathematics Laboratory, N.Z.D.S.I.R., Wellington, and to thank N.Z.D.S.I.R. for permission to publish this short paper.

REFERENCES

BRAVAIS, A. (1846). Analyse mathématique sur les probabilités des erreurs de situation d'un point. *Mém. prés. Acad. Sci., Paris*, **9**, 255–332.

BUNIMOVITCH, V. I. (1949). *J. Tech. Phys.* (U.S.S.R.), **19**. Cited (p. 1361) by Magness, T.A. (1954). Spectral response of a quadratic device to non-Gaussian noise. *J. Appl. Phys.* **25**, 1357–65.

CRAMÉR, H. (1937). *Random Variables and Probability Distributions*, Camb. Math. Tract, no. 36.

PEARSON, K. (1896). Mathematical contributions to the theory of evolution. III. *Phil. Trans.* A, **187**, 253–318.

RICE, S. O. (1944). Mathematical analysis of random noise. *Bell Syst. Tech. J.* **23**, 282–332.

RICE, S. O. (1945). Mathematical analysis of random noise (concluded). *Bell Syst. Tech. J.* **24**, 46–156.

Stationarity conditions for stochastic processes of the autoregressive and moving-average type

BY J. WISE

University of Birmingham

INTRODUCTION AND SUMMARY*

The criterion for the stationarity of a process of the autoregressive and moving-average types has been given by Wold (1954) in the form that the roots of a specified equation shall lie inside the unit circle. A generalization of this result has been given by Doob (1953), who showed that a stationary stochastic process of the autoregressive or moving-average type has a rational spectral density function $g(z)$, where

$$g(z) = \frac{A(z)\,A(\bar{z})}{B(z)\,B(\bar{z})}, \tag{1}$$

$A(z)$ and $B(z)$ being polynomials in z with no common factors, the roots of the equation

$$B(z) = 0 \tag{2}$$

all lying inside the unit circle, and no root of the equation

$$A(z) = 0 \tag{3}$$

lying outside the unit circle.

In this paper these results due to Wold and Doob are converted into a set of criteria which are determined directly from the coefficients of the polynomials concerned. No intermediate evaluation of the roots of the equation is necessary. Thus a set of necessary and sufficient conditions for the stationarity of a stochastic process of the autoregressive or moving-average type is derived explicitly in terms of the coefficients of the process. These criteria are expressed in a determinantal form permitting systematic treatment in the general case. All the elements of the determinants involved, for polynomials up to (and including) the fourth degree, are given, and these determinants are given in expanded form. In addition, a set of necessary (but not sufficient) conditions are given for stationarity, which do not involve determinants and which are of a very simple algebraic form.

DERIVATION OF THE CRITERIA

Let us consider the autoregressive moving-average process

$$x_t + \alpha_1 x_{t-1} + \ldots + \alpha_k x_{t-k} = \epsilon_t + \beta_1 \epsilon_{t-1} + \ldots + \beta_h \epsilon_{t-h}. \tag{4}$$

* This paper is based on material contained in Wise (1955*a*).

We may assume, without loss of generality, that the polynomials

$$\left.\begin{aligned} z^k+\alpha_1 z^{k-1}+\ldots+\alpha_k \\ z^h+\beta_1 z^{h-1}+\ldots+\beta_h \end{aligned}\right\} \quad (5)$$

and
have no common factor.

Then, if the process (4) is stationary, all the roots of the equation

$$z^k+\alpha_1 z^{k-1}+\ldots+\alpha_k=0 \quad (6)$$

lie inside the unit circle, and none of the roots of the equation

$$z^h+\beta_1 z^{h-1}+\ldots+\beta_h=0 \quad (7)$$

lies outside the unit circle if the process is non-circular, while all the roots lie inside the unit circle if the process is circular.* Let us consider the polynomial equation

$$x^k+\alpha_1 x^{k-1}+\ldots+\alpha_k=0. \quad (8)$$

The roots of this equation are given by the reduction of the polynomial to real or complex linear factors such that

$$(x-x_1)(x-x_2)\ldots(x-x_k)=0. \quad (9)$$

We now require the necessary and sufficient conditions for the satisfaction of the inequalities

$$|x_j|<1 \quad (j=1,2,3,\ldots,k). \quad (10)$$

Put

$$y=\frac{x+1}{x-1} \quad \text{and} \quad x=a+ib \quad (11)$$

(where a and b are real). We find

$$y=\frac{a^2+b^2-1}{(a-1)^2+b^2}-\frac{2ib}{(a-1)^2+b^2} \quad (12)$$

$$=R(y)+iC(y), \quad (13)$$

where $R(y)$ denotes the real part and $iC(y)$ denotes the complex part of y. We have

$$R(y)=\frac{a^2+b^2-1}{(a-1)^2+b^2}, \quad (14)$$

$$C(y)=-\frac{2b}{(a-1)^2+b^2}, \quad (15)$$

$$|x|=+\sqrt{(a^2+b^2)}, \quad (16)$$

$$(a-1)^2+b^2\geqslant 0. \quad (17)$$

Thus from (14), (16) and (17) it follows that it is necessary and sufficient for $|x|<1$ that

$$R(y)<0. \quad (18)$$

From (11) we obtain

$$x=\frac{y+1}{y-1}, \quad (19)$$

so that (10) is satisfied, if, and only if

$$R(y_j)<0 \quad (j=1,2,\ldots,k), \quad (20)$$

where $y_j\,(j=1,2,\ldots,k)$ are the roots of the equation

$$(y+1)^k+\alpha_1(y+1)^{k-1}(y-1)+\ldots+\alpha_k(y-1)^k=0, \quad (21)$$

which is obtained by substituting for x from (19) into (8). Expressing (21) in ascending powers of y we obtain

$$p_0+p_1y+p_2y^2+\ldots+p_ky^k=0, \quad (22)$$

where

$$p_r=\sum_{j=0}^{k}\alpha_j c_{rj}, \quad (23)$$

$\alpha_0=1$ and c_{rj} is the coefficient of y^r in the expression $(y+1)^{k-j}(y-1)^j$.

The values of the coefficients c_{rj} are tabulated below for $k=1,2,3,4$.

* See Wise (1955*b*) for a discussion of circular and non-circular processes of the autoregressive and moving-average type.

From these the coefficients $p_1, \ldots, p_k$ are tabulated for $k = 1, 2, 3, 4$, expressed in terms of the coefficients $\alpha_1, \alpha_2, \alpha_3, \alpha_4$, on the condition p_0 having been set equal to unity.

Necessary and sufficient conditions for (20) to be satisfied, expressed as criteria involving only the coefficients in equation (22), have been derived by Routh (1930) and are given in convenient form below.

The criteria are obtained as follows. We form the matrix P, where

$$P = \begin{bmatrix} p_1 & p_3 & p_5 & p_7 & \cdots \\ p_0 & p_2 & p_4 & p_6 & \cdots \\ 0 & p_1 & p_3 & p_5 & \cdots \\ 0 & p_0 & p_2 & p_4 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}. \tag{24}$$

The Routh conditions then state that (20) is satisfied, and thus (10) is satisfied, if, and only if, the following inequalities are satisfied (p_0 having been put equal to unity):

$$\left.\begin{aligned} & P_j = p_j > 0 \quad (j = 1 \text{ and } j = k) \\ & P_2 = \begin{vmatrix} p_1 & p_3 \\ p_0 & p_2 \end{vmatrix} > 0 \\ & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ & P_r > 0 \qquad (r = 1, 2, \ldots, k-1), \end{aligned}\right\} \tag{25}$$

where $P_r\,(r = 1, 2, \ldots, k-1)$ is the principal minor of P formed from the first r rows and the first r columns. The principal minor formed from the first k rows and columns equals $p_k P_{k-1}$, and since both this and P_{k-1} must be positive, this implies that p_k is positive. A set of necessary, but not sufficient, conditions for (10) is given by the inequalities

$$p_r > 0 \quad (r = 1, 2, \ldots, k). \tag{26}$$

Expressions for $p_r\,(r = 1, 2, \ldots, k)$ are given below, in terms of the coefficients of equation (8), for $k = 1, 2, 3, 4$. This covers nearly all cases likely to occur in practice.* From these values the criteria given by (25) may be obtained directly. These criteria are given below, in expanded form, for $k = 1, 2, 3$ and 4. The expansions of the determinants in terms of the coefficients of equation (8) become unwieldy for values of k larger than 4. For cases in which k exceeds 4, it is more convenient to work with the P criteria in determinantal form. It is worth noting, furthermore, that the computational labour involved in evaluating the P criteria is small relative to the labour required to evaluate the roots of (8) numerically.

Values of c_{rj}

$k = 1$

j \ r	0	1
0	1	1
1	−1	1

$k = 2$

j \ r	0	1	2
0	1	2	1
1	−1	0	1
2	1	−2	1

$k = 3$

j \ r	0	1	2	3
0	1	3	3	1
1	−1	−1	1	1
2	1	−1	−1	1
3	−1	3	−3	1

$k = 4$

j \ r	0	1	2	3	4
0	1	4	6	4	1
1	−1	−2	0	2	1
2	1	0	−2	0	1
3	−1	2	0	−2	1
4	1	−4	6	−4	1

* For more extensive tabulations for $k = 1, 2, \ldots, 10$, see Wise (1955*a*).

Tables of the p-coefficients

$k = 1$:

$$p_1 = \frac{1+\alpha_1}{1-\alpha_1}, \quad p_2 = 0.$$

$k = 2$:

$$p_1 = \frac{2(1-\alpha_2)}{(1-\alpha_1+\alpha_2)}, \quad p_2 = \frac{1+\alpha_1+\alpha_2}{1-\alpha_1+\alpha_2}.$$

$k = 3$:

$$p_1 = \frac{(3-\alpha_1-\alpha_2+3\alpha_3)}{(1-\alpha_1+\alpha_2-\alpha_3)},$$

$$p_2 = \frac{(3+\alpha_1-\alpha_2-3\alpha_3)}{(1-\alpha_1+\alpha_2-\alpha_3)},$$

$$p_3 = \frac{(1+\alpha_1+\alpha_2+\alpha_3)}{(1-\alpha_1+\alpha_2-\alpha_3)}.$$

$k = 4$:

$$p_1 = \frac{2(2-\alpha_1+\alpha_3-2\alpha_4)}{(1-\alpha_1+\alpha_2-\alpha_3+\alpha_4)},$$

$$p_2 = \frac{2(3-\alpha_2+3\alpha_4)}{(1-\alpha_1+\alpha_2-\alpha_3+\alpha_4)},$$

$$p_3 = \frac{2(2+\alpha_1-\alpha_3-2\alpha_4)}{(1-\alpha_1+\alpha_2-\alpha_3+\alpha_4)},$$

$$p_4 = \frac{(1+\alpha_1+\alpha_2+\alpha_3+\alpha_4)}{(1-\alpha_1+\alpha_2-\alpha_3+\alpha_4)}.$$

Tables of the P-criteria

$k = 1$:

$$P_1 = p_1$$
$$= \frac{1+\alpha_1}{1-\alpha_1} > 0.$$

$k = 2$:

$$P_1 = p_1$$
$$= \frac{2(1-\alpha_2)}{(1-\alpha_1+\alpha_2)} > 0,$$
$$P_2 = p_2$$
$$= \frac{(1+\alpha_1+\alpha_2)}{(1-\alpha_1+\alpha_2)} > 0.$$

$k = 3$:

$$P_1 = p_1$$
$$= \frac{(3-\alpha_1-\alpha_2+3\alpha_3)}{(1-\alpha_1+\alpha_2-\alpha_3)} > 0,$$
$$P_2 = p_1 p_2 - p_3$$
$$= \frac{(3-\alpha_1-\alpha_2+3\alpha_3)(3+\alpha_1-\alpha_2-3\alpha_3)}{(1-\alpha_1+\alpha_2-\alpha_3)^2} - \frac{(1+\alpha_1+\alpha_2+\alpha_3)}{(1-\alpha_1+\alpha_2-\alpha_3)} > 0$$
$$P_3 = p_3$$
$$= \frac{(1+\alpha_1+\alpha_2+\alpha_3)}{(1-\alpha_1+\alpha_2-\alpha_3)} > 0.$$

$k = 4$:

$$P_1 = p_1 = \frac{2(2-\alpha_1+\alpha_3-2\alpha_4)}{(1-\alpha_1+\alpha_2-\alpha_3+\alpha_4)} > 0,$$

$$P_2 = p_1p_2 - p_3 = \frac{4(2-\alpha_1+\alpha_3-2\alpha_4)(3-\alpha_2+\alpha_4)}{(1-\alpha_1+\alpha_2-\alpha_3+\alpha_4)^2} - \frac{2(2+\alpha_1-\alpha_3-2\alpha_4)}{(1-\alpha_1+\alpha_2-\alpha_3+\alpha_4)} > 0,$$

$$P_3 = p_1p_2p_3 - p_1^2p_4 - p_3^2 = \frac{8(2-\alpha_1+\alpha_3-2\alpha_4)(3-\alpha_2+3\alpha_4)(2+\alpha_1-\alpha_3-2\alpha_4)}{(1-\alpha_1+\alpha_2-\alpha_3+\alpha_4)^3} - \frac{4(2-\alpha_1+\alpha_3-2\alpha_4)^2(1+\alpha_1+\alpha_2+\alpha_3+\alpha_4)}{(1-\alpha_1+\alpha_2-\alpha_3+\alpha_4)^3} - \frac{4(2+\alpha_1-\alpha_3-2\alpha_4)^2}{(1-\alpha_1+\alpha_2-\alpha_3+\alpha_4)^2} > 0,$$

$$P_4 = \frac{(1+\alpha_1+\alpha_2+\alpha_3+\alpha_4)}{(1-\alpha_1+\alpha_2-\alpha_3+\alpha_4)} > 0.$$

REFERENCES

Doob, J. L. (1953). *Stochastic Processes.* New York: John Wiley and Sons.

Routh, E. J. (1930). *Dynamics of a System of Rigid Bodies.* London: Macmillan.

Wise, J. (1955*a*). Ph.D. Thesis (unpublished), University of London.

Wise, J. (1955*b*). The autocorrelation function and the spectral density function. *Biometrika*, **42**, 151.

Wold, H. (1954). *A Study in the Analysis of Stationary Time Series*, Uppsala: Almqvist and Wicksell.

Some properties of an angular transformation for the correlation coefficient

By B. I. HARLEY

University College London

1. Sheppard (1899), in an early paper in which he discussed the association between two variables, used a quantity which is equivalent to the arc sine of the correlation coefficient. As far as the writer knows, the properties of such a quantity have not been discussed, although Jenkins (1954), following suggestions put forward by others, did investigate the arc sine of the serial correlation coefficient. It is the purpose of this present note to take this early work of Sheppard a stage further and in particular to show that if r is the product-moment correlation in the sample, and ρ that in the normal population, then

$$\mathscr{E}(\sin^{-1} r) = \sin^{-1}\rho.$$

2. Let

$$y = \sin^{-1} r,$$

and expand in a Taylor series about the point $r = \rho$. Using the moments of r expressed in the form of a series, we have on taking expectations and retaining terms of order n^{-3} that

$$\mathscr{E}(y) = \sin^{-1}\rho, \tag{1}$$

$$\kappa_2(y) = \frac{(1-\rho^2)}{n}\left\{1+\frac{1}{n}(2+\rho^2)+\frac{1}{3n^2}(11+8\rho^2+8\rho^4)\right\}, \tag{2}$$

$$\kappa_3(y) = -\frac{3\rho(1-\rho^2)^{\frac{3}{2}}}{n^2}\left\{1+\frac{1}{n}(3+4\rho^2)\right\}, \tag{3}$$

$$\kappa_4(y) = \frac{2(1-\rho^2)^2}{n^3}\{10\rho^2-1\}, \tag{4}$$

from which we deduce

$$\beta_1(y) = \frac{9\rho^2}{n}\left\{1+\frac{5\rho^2}{n}-\frac{2}{n^2}(7+13\rho^2+5\rho^4)\right\}+\ldots, \tag{5}$$

$$\beta_2(y) = 3+\frac{2}{n}(10\rho^2-1)\left\{1-\frac{2}{n}(2+\rho^2)+\ldots\right\}. \tag{6}$$

For large n the quantity $\sin^{-1}r$ is therefore approximately normally distributed.

3. R. A. Fisher (1921) gave the normalizing transformation

$$z = \tfrac{1}{2}\log_e\frac{1+r}{1-r} = \tanh^{-1}r,$$

for which (making use of the correction given by A. K. Gayen (1951)) we have the moments

$$\mathscr{E}(z) = \tanh^{-1}\rho+\frac{\rho}{2(n-1)}\left\{1+\frac{5+\rho^2}{4(n-1)}+\ldots\right\}, \tag{7}$$

$$\kappa_2(z) = \frac{1}{(n-1)}\left\{1+\frac{4-\rho^2}{2(n-1)}+\frac{22-6\rho^2-3\rho^4}{6(n-1)^2}+\ldots\right\}, \tag{8}$$

and the β coefficients

$$\beta_1(z) = \frac{\rho^6}{(n-1)^3}+\ldots, \tag{9}$$

$$\beta_2(z) = 3+\frac{2}{n-1}+\frac{4+2\rho^2-3\rho^4}{(n-1)^2}+\ldots. \tag{10}$$

Table 1. *Comparison of the values of β_1 and β_2 for the distributions r, z and y*

$n = 20$

ρ	$\beta_1(r)$	$\beta_1(z)$	$\beta_1(y)$	$\beta_2(r)$	$\beta_2(z)$	$\beta_2(y)$
0·1	0·0148	$10^{-9}\times 0{\cdot}1458$	0·0044	2·7406	3·1164	2·9281
0·2	0·0600	$10^{-8}\times 0{\cdot}9331$	0·0175	2·8213	3·1166	2·9522
0·3	0·1386	$10^{-6}\times 0{\cdot}1063$	0·0397	2·9623	3·1168	2·9921
0·4	0·2557	$10^{-6}\times 0{\cdot}5972$	0·0716	3·1749	3·1170	3·0470
0·5	0·4202	$10^{-5}\times 0{\cdot}2278$	0·1136	3·4783	3·1172	3·1162
0·6	0·6464	$10^{-5}\times 0{\cdot}6802$	0·1666	3·9055	3·1173	3·1986
0·7	0·9584	$10^{-4}\times 0{\cdot}1715$	0·2314	4·5131	3·1171	3·2929
0·8	1·4001	$10^{-4}\times 0{\cdot}3822$	0·3091	5·4154	3·1165	3·3974
0·9	2·0603	$10^{-4}\times 0{\cdot}7748$	0·4004	6·8681	3·1154	3·5105

Apart from its normalizing properties, the great practical advantage of the z transformation is that it gives a variance which, to order $(n-1)^{-1}$, is independent of ρ. The y transformation does not have this property but it provides a much neater value for the expectation, which is also independent of n. Without going into the question of the application of the results, it is proposed to make some numerical comparisons of the properties of the $\sin^{-1}r$ and $\tanh^{-1}r$ distributions.

A comparison of the values of $\beta_1(z)$ and $\beta_2(z)$ with $\beta_1(y)$ and $\beta_2(y)$ for $n = 20$ and $\rho = 0{\cdot}1\,(0{\cdot}1)\,0{\cdot}9$ are given in Table 1. The values of $\beta_1(r)$ and $\beta_2(r)$, as given in *Tables for Statisticians and Biometricians*, vol. 2, are given for completeness. It will be noted that $\beta_1(z)$ is much smaller than $\beta_1(y)$ and this will be true for all values of ρ for $n \geqslant 5$. $\beta_2(z)$ is more independent of ρ than $\beta_2(y)$ which follows more closely the values for the r distribution. From an examination of the results it seems unlikely that the y transformation will have a better normalizing effect than the z transformation.

4. To compare the probability integral of r obtained by using the $\sin^{-1}r$ transformation with the values obtained by using the $\tanh^{-1}r$ transformation we take as a first approximation

$$y = \sin^{-1}r, \quad \mathscr{E}(y) = \sin^{-1}\rho, \quad \kappa_2(y) = (1-\rho^2)/(n-2),$$

the divisor, $n-2$, for the variance giving a good approximation to the expansion (2) when ρ is small. Using this mean and variance, a normal curve can be fitted and the probability integral of r obtained for various values of n and ρ. As a second approximation a normal curve can be fitted using the values for the mean and variance given in equations (1) and (2) respectively.

Table 2. *Comparison of exact and approximate values of the probability integral of r*

(i) $n = 25$, $\rho = 0{\cdot}2$

r	Exact value	From z 1st approx.	From z 2nd approx.	From y 1st approx.	From y 2nd approx.
−0·2	0·02 748	0·02 860	0·02 716	0·02 435	0·02 439
−0·1	0·07 379	0·07 758	0·07 444	0·06 999	0·07 005
0	0·16 364	0·17 083	0·16 542	0·16 217	0·16 225
0·1	0·30 630	0·31 551	0·30 807	0·31 020	0·31 025
0·2	0·49 168	0·50 000	0·49 179	0·50 000	0·50 000
0·3	0·68 635	0·69 177	0·68 466	0·69 350	0·69 344
0·4	0·84 700	0·84 994	0·84 533	0·84 818	0·84 810
0·5	0·94 615	0·94 798	0·94 592	0·94 263	0·94 258
0·6	0·98 820	0·98 928	0·98 875	0·98 477	0·98 475
0·7	0·99 877	0·99 909	0·99 903	0·99 752	0·99 751

(ii) $n = 25$, $\rho = 0{\cdot}6$

r	Exact value	From z 1st approx.	From z 2nd approx.	From y 1st approx.	From y 2nd approx.
0·2	0·00 934	0·01 072	0·00 884	0·00 402	0·00 424
0·3	0·03 097	0·03 598	0·03 078	0·02 112	0·02 186
0·4	0·09 046	0·10 311	0·09 147	0·08 216	0·08 365
0·5	0·22 771	0·24 994	0·22 971	0·23 614	0·23 769
0·6	0·47 500	0·50 000	0·47 521	0·50 000	0·50 000
0·7	0·77 782	0·79 299	0·77 584	0·78 544	0·78 381
0·75	0·89 652	0·90 531	0·89 543	0·88 995	0·88 833
0·8	0·96 741	0·97 140	0·96 769	0·95 555	0·95 442
0·85	0·99 469	0·99 586	0·99 520	0·98 722	0·98 670

Similarly, if $z = \tanh^{-1} r$, we take as a first approximation

$$\mathscr{E}(z) = \tanh^{-1}\rho, \quad \kappa_2(z) = 1/(n-3),$$

and then, as a second approximation, take the values for the mean and variance given in equations (7) and (8) respectively, and in each case fit a normal curve with the appropriate mean and variance. The values of the probability integral of r obtained by these approximate methods, together with the exact values as given in F. N. David's tables (1938), for the case where $n = 25$, $\rho = 0{\cdot}2$, $0{\cdot}6$ are given in Table 2.

From an examination of the results it is seen that in the case of the z transformation, the inclusion of further terms of the expansions for the mean and variance of z increases the accuracy of the approximations considerably, and the values using the second approximation for z are closer to the exact values than any of the other approximations. Increasing the number of terms of the expansion for the variance in the case of the y distribution makes very little difference, since the skewness of the y distribution prevents it from being as good as the z distribution.

5. Previously (1954) we have given the Cornish–Fisher (1937) form of the Edgeworth expansion fitted to the distribution of $\tanh^{-1} r$, using the values of the first four cumulants to the full known accuracy of their expansions. For the purposes of comparison we use the same series expansion to approximate to the distribution of $\sin^{-1} r$, using the cumulants as given in equations (1)–(4). The probability integral values for the cases $n = 16$, $\rho = 0{\cdot}2$ and $n = 25$, $\rho = 0{\cdot}9$, are given in Table 3.

Table 3. *Comparison of exact and approximate values of the probability integral of r*

(i) $n = 16$, $\rho = 0{\cdot}2$, $\beta_1(y) = 0{\cdot}0215$, $\beta_2(y) = 2{\cdot}9441$

r	Exact value	Edgeworth expansion for y
$-0{\cdot}2$	0·06 711	0·06 722
$-0{\cdot}1$	0·12 868	0·12 953
0	0·22 076	0·22 202
0·1	0·34 349	0·34 367
0·2	0·48 934	0·48 806
0·3	0·64 273	0·64 123
0·4	0·78 316	0·78 297
0·5	0·89 186	0·89 285
0·6	0·95 947	0·96 026
0·7	0·99 034	0·99 035

(ii) $n = 25$, $\rho = 0{\cdot}9$, $\beta_1(y) = 0{\cdot}0319$, $\beta_2(y) = 3{\cdot}4403$, $\beta_1(z) = 10^{-4} \times 0{\cdot}3844$, $\beta_2(z) = 3{\cdot}0897$

r	Exact value	Edgeworth expansion for y	Edgeworth expansion for z
0·75	0·00 743	0·00 684	0·00 742
0·82	0·05 574	0·05 834	0·05 578
0·87	0·22 387	0·22 896	0·22 379
0·90	0·46 244	0·45 513	0·46 247
0·93	0·78 645	0·78 143	0·78 661
0·95	0·94 612	0·95 047	0·94 612
0·965	0·99 263	0·99 660	0·99 264

This method of approximation is unlikely to give as good values from the y transformation as from the z, because the former has not secured so close an approach to normality as the latter. The y transformation may be useful as a quick check, however, for since the expectation of y is known exactly a number of the terms in the Edgeworth expansion is zero, and it is simple to compute.

6. From the moments of § 2 it appears that

$$\mathscr{E}(\sin^{-1} r) = \sin^{-1} \rho$$

up to terms of order n^{-3}. The identity holds in fact for any order of n^{-1}.

The probability distribution of r, the correlation coefficient, calculated from a sample of size n, randomly drawn from a normal bivariate population with correlation coefficient ρ is known to be

$$p(r|n,\rho) = \frac{(1-\rho^2)^{\frac{1}{2}(n-1)}}{\pi(n-3)!}(1-r^2)^{\frac{1}{2}(n-4)}\frac{d^{n-2}}{d(r\rho)^{n-2}}\left(\frac{\cos^{-1}(-r\rho)}{(1-r^2\rho^2)^{\frac{1}{2}}}\right). \tag{11}$$

Since it is a probability distribution,

$$\int_{-1}^{+1}(1-r^2)^{\frac{1}{2}(n-4)}\frac{d^{n-2}}{d(r\rho)^{n-2}}\left(\frac{\cos^{-1}(-r\rho)}{(1-r^2\rho^2)^{\frac{1}{2}}}\right)dr=\frac{\pi(n-3)!}{(1-\rho^2)^{\frac{1}{2}(n-1)}}=C_n(\rho).$$

Denoting $\dfrac{d^{n-2}}{d(r\rho)^{n-2}}\left(\dfrac{\cos^{-1}(-r\rho)}{(1-r^2\rho^2)^{\frac{1}{2}}}\right)$ by $f^{n-2}(r\rho)$, and rearranging the integral, we have

$$\int_{-1}^{+1}\frac{1}{(1-r^2)^{\frac{1}{2}}}(1-r^2)^{\frac{1}{2}(n-3)}f^{n-2}(r\rho)=C_n(\rho). \tag{12}$$

Integrating by parts we have

$$C_n(\rho)=[\sin^{-1}r(1-r^2)^{\frac{1}{2}(n-3)}f^{n-2}(r\rho)]_{-1}^{+1}-\int_{-1}^{+1}\sin^{-1}r\frac{\partial}{\partial r}\{(1-r^2)^{\frac{1}{2}(n-3)}f^{n-2}(r\rho)\}\,dr.$$

For $n\geqslant 3$, $[\sin^{-1}r(1-r^2)^{\frac{1}{2}(n-3)}f^{n-2}(r\rho)]_{-1}^{+1}=0$, since $f^{n-2}(r\rho)$ is finite and non-zero at the limits ± 1. Therefore

$$C_n(\rho)=-\int_{-1}^{+1}\sin^{-1}r(1-r^2)^{\frac{1}{2}(n-3)}\rho\frac{d}{d(r\rho)}f^{n-2}(r\rho)\,dr+(n-3)\int_{-1}^{+1}r\sin^{-1}r(1-r^2)^{\frac{1}{2}(n-5)}f^{n-2}(r\rho)\,dr,$$

and hence

$$C_n(\rho)=-\rho\int_{-1}^{+1}\sin^{-1}r(1-r^2)^{\frac{1}{2}(n-3)}f^{n-1}(r\rho)\,dr+(n-3)\int_{-1}^{+1}\frac{\partial}{\partial\rho}\{\sin^{-1}r\,p(r|(n-1),\rho)C_{n-1}(\rho)\}\,dr. \tag{13}$$

Denoting the expectation of $\sin^{-1}r$, when r is calculated from a sample of size n, by $\mathscr{E}_n(\sin^{-1}r)$, we have

$$C_n(\rho)=-\rho C_{n+1}(\rho)\,\mathscr{E}_{n+1}(\sin^{-1}r)+(n-3)\left[C_{n-1}(\rho)\frac{\partial}{\partial\rho}\mathscr{E}_{n-1}(\sin^{-1}r)+\mathscr{E}_{n-1}(\sin^{-1}r)\frac{d}{d\rho}C_{n-1}(\rho)\right]. \tag{14}$$

Substituting the values of the C's and rearranging we have

$$\frac{1}{(1-\rho^2)^{\frac{1}{2}}}-\frac{\partial}{\partial\rho}\mathscr{E}_{n-1}(\sin^{-1}r)=\frac{\rho(n-2)}{(1-\rho^2)}[\mathscr{E}_{n-1}(\sin^{-1}r)-\mathscr{E}_{n+1}(\sin^{-1}r)]. \tag{15}$$

If $\mathscr{E}_{n-1}(\sin^{-1}r)=\sin^{-1}\rho$, since we have assumed $n>2$, it follows that

$$\mathscr{E}_{n+1}(\sin^{-1}r)=\mathscr{E}_{n-1}(\sin^{-1}r)=\sin^{-1}\rho.$$

From (12) when $n=3$ we have

$$\begin{aligned}C_3(\rho)&=\int_{-1}^{+1}(1-r^2)^{-\frac{1}{2}}f^1(r\rho)\,dr\\&=\left[\sin^{-1}r\left\{\frac{1}{1-r^2\rho^2}+\frac{\rho r\cos^{-1}(-r\rho)}{(1-r^2\rho^2)^{\frac{1}{2}}}\right\}\right]_{-1}^{+1}-\rho C_4(\rho)\,\mathscr{E}_4(\sin^{-1}r).\end{aligned}$$

Hence

$$\frac{\pi}{1-\rho^2}=\frac{\pi}{1-\rho^2}+\frac{\pi\rho\sin^{-1}\rho}{(1-\rho^2)^{\frac{1}{2}}}-\frac{\rho\pi}{(1-\rho^2)^{\frac{1}{2}}}\mathscr{E}_4(\sin^{-1}r),$$

and therefore

$$\mathscr{E}_4(\sin^{-1}r)=\sin^{-1}\rho. \tag{16}$$

From equation (15) it follows that $\mathscr{E}_6(\sin^{-1}r)=\mathscr{E}_4(\sin^{-1}r)=\sin^{-1}\rho$ and, in fact, $\mathscr{E}_{2n}(\sin^{-1}r)=\sin^{-1}\rho$ for $n=2,3,\ldots$. It has not been possible to show by the same method that $\mathscr{E}_n(\sin^{-1}r)$ for odd sample sizes is equal to $\sin^{-1}\rho$. However, if we assume

$$\mathscr{E}_n(\sin^{-1}r)=\sin^{-1}\rho+\frac{A_4(\rho)}{n^4}+\frac{A_5(\rho)}{n^5}+\ldots$$

for all values of n, where $A_s(\rho)$ for $s=4,5,\ldots$ are functions of ρ, then substituting in equation (15) we have

$$\begin{aligned}0=\frac{\rho}{1-\rho^2}(n-2)&\left[A_4(\rho)\left\{\frac{1}{(n-1)^4}+\frac{1}{(n+1)^4}\right\}+A_5(\rho)\left\{\frac{1}{(n-1)^5}+\frac{1}{(n+1)^5}\right\}+\ldots\right]\\&+\left[\frac{A_4'(\rho)}{(n-1)^4}+\frac{A_5'(\rho)}{(n-1)^5}+\ldots\right],\end{aligned}$$

where

$$A_s'(\rho)=\frac{d}{d\rho}A_s(\rho)\quad\text{for } s=4,5,\ldots.$$

Equating coefficients of n^{-3} we find that $A_4(\rho)=0$. Similarly, we find that $0=A_5(\rho)=A_6(\rho)=\ldots$. It seems, therefore, reasonable to assume that $\mathscr{E}_n(\sin^{-1}r)=\sin^{-1}\rho$ for all values of n.

7. In his paper 'New light on the correlation coefficient and its transforms', H. Hotelling (1953, p. 221) states as a theorem:

'No function $\psi(r)$, independent of ρ and n, satisfying the conditions for expansibility in a Taylor series through terms of fourth order, and termwise integrable to give the expectation in a series of powers of n^{-1}, exists such that $\mathscr{E}(\psi(r)) = \psi(\rho)$.'

We think, however, that this conclusion is due to a mistake, for if the substitution which Hotelling refers to in the penultimate line of his p. 220 is made, we find that the coefficient of n^{-2} in his earlier expansion for $\mathscr{E}(\psi(r)) - \psi(\rho)$ vanishes identically for all values of ρ.

I wish to thank Dr F. N. David and Dr D. E. Barton for the helpful suggestions they made during the preparation of this paper.

REFERENCES

Cornish, E. A. & Fisher, R. A. (1937). *Rev. Inst. int. Statist., Paris*, **4**, 307.
David, F. N. (1938). *Tables of the Ordinates and Probability Integral of the Distribution of the Correlation Coefficient in Small Samples*. Cambridge University Press.
Fisher, R. A. (1921). *Metron*, **1**, no. 4, p. 1.
Gayen, A. K. (1951). *Biometrika*, **38**, 219.
Harley, B. I. (1954). *Biometrika*, **41**, 278.
Hotelling, H. (1953). *J. R. Statist. Soc.* B, **15**, 193.
Jenkins, G. M. (1954). *Biometrika*, **41**, 261.
Pearson, K. (1931). *Tables for Statisticians and Biometricians*, **2**. Cambridge University Press.
Sheppard, W. F. (1899). *Phil. Trans.* A, **192**, 101.

Note on the moment-problem for unimodal distributions when one or both terminals are known

By C. L. MALLOWS

University College London

1. Several mathematical models have been proposed in attempts to represent biological data relating, for example, to counts of species in randomly chosen regions. Anscombe (1950) gives a discussion, and compares eight different two-parameter distributions. He remarks that of these distributions, some have only one mode, some one or two, while some may have any number from one upwards. This progression is reflected in the behaviour of the third and fourth cumulants; the number of modes appears to increase with a decrease in both skewness and leptokurtosis.

It is of interest to determine whether in fact this must be so; more specifically, to determine the 'unimodality boundary' for such distributions, i.e. a boundary defining values of the lower moments such that any distribution having these moments must of necessity be multimodal.

2. The necessary and sufficient conditions which the sequence of moments $\mu_0, \mu_1, \ldots, \mu_n$ must satisfy in order that a (cumulative) distribution function $\phi(x)$ having these moments may exist are known in the situations where x

(a) may take any value between $-\infty$ and $+\infty$;
(b) is restricted to positive values;
(c) is restricted to lie in a finite interval (t_1, t_2) (see, for example, Shohat & Tamarkin (1943)).

By means of a transformation Johnson & Rogers (1951) reduced the case (a), where the additional restriction is applied that the distribution is to be unimodal, to the case (a) without restriction. It is now shown that this transformation can also be applied to the cases (b) and (c) above.

These results apply immediately only to continuous distributions; however, we can obtain results for the discrete case by considering the distribution of the sum of the discrete variable and an independent rectangular variable. This will have a 'histogram' distribution which may be regarded as continuous.

3. The conditions referred to in §2 are as follows (moments about zero throughout)*:

Case (*a*): $t_1 = -\infty$, $t_2 = +\infty$. The determinants

$$\Delta_r(\mu) = |\,\mu_{i+j}\,|_{i,j=0}^{r} \quad (r = 0, 1, \ldots, [\tfrac{1}{2}n])$$

are positive.

Case (*b*): $t_1 = 0$, $t_2 = +\infty$. The determinants

$$\Delta_r(\mu) \quad (r = 0, 1, \ldots, [\tfrac{1}{2}n]), \qquad \Delta_r'(\mu) = |\,\mu_{i+j+1}\,|_{i,j=0}^{r} \quad (r = 0, 1, \ldots, [\tfrac{1}{2}(n-1)])$$

are positive.

Case (*c*): $t_1 = -1$, $t_2 = +1$. If n is even $= 2m$, the determinants

$$\Delta_r(\mu), \quad \Delta_r''(\mu) = |\,\mu_{i+j} - \mu_{i+j+2}\,|_{i,j=0}^{r-1} \quad (r = 0, 1, \ldots, m)$$

are positive; if $n = 2m+1$, the determinants

$$\Delta_r'''(\mu) = |\,\mu_{i+j} + \mu_{i+j+1}\,|_{i,j=0}^{r}, \quad \Delta_r^{\mathrm{iv}}(\mu) = |\,\mu_{i+j} - \mu_{i+j+1}\,|_{i,j=0}^{r} \quad (r = 0, 1, \ldots, m)$$

are positive.

4. The transformation used by Johnson & Rogers in the case (*a*) is as follows. Suppose $\phi''(x)$ exists for all x, and

$$\phi''(x) \begin{cases} \geqslant 0 & (x < \beta), \\ = 0 & (x = \beta), \\ \leqslant 0 & (x > \beta). \end{cases}$$

Then the distribution is unimodal, and has a mode at β. Note that we may have $\phi''(x)$ vanishing over an interval including β, so that there is no unique mode. Any unimodal distribution (continuous, but whose second derivative need not exist; e.g. a 'histogram' distribution) can be approximated as closely as desired by such a $\phi(x)$.

Consider the function $\psi(x)$ given by

$$\psi'(x) = (\beta - x)\,\phi''(x), \quad \psi(-\infty) = 0.$$

Then $\psi(x)$ is a distribution function with moments

$$\nu_r = \int x^r d\psi(x) = (r+1)\,\mu_r - \beta r \mu_{r-1},$$

and so the ν's must satisfy the conditions of §3 (*a*). We thus obtain limits for β for the given moments $\mu_0, \mu_1, \ldots, \mu_n$, and may derive the required conditions on the moments themselves.

The determinant $\Delta_r(\nu)$ can be manipulated into the following form:

$$\Delta_r(\nu) = \begin{vmatrix} 1 & \beta & \beta^2 & \ldots & \beta^{r+1} \\ 0 & \mu_0 & 2\mu_1 & \ldots & (r+1)\,\mu_r \\ \mu_0 & 2\mu_1 & 3\mu_2 & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ r\mu_{r-1} & \ldots & \ldots & \ldots & (2r+1)\,\mu_{2r} \end{vmatrix},$$

which will be denoted by $\Delta_r(\nu) = |\,\beta^j;\ (i+j-1)\,\mu_{i+j-2}\,|_{i,j=0}^{r+1}$.

5. The same transformation can be applied in the cases (*b*) and (*c*) above. We obtain the following results, where the moments are about zero throughout.

For a unimodal solution of the moment problem to exist with a mode at β, it is necessary and sufficient that:

Case (*b*). Terminals $t_1 = 0$, $t_2 = +\infty$. The determinants

$$\Delta_r(\nu) \quad (r = 0, 1, \ldots, [\tfrac{1}{2}n]), \qquad \Delta_r'(\nu) = |\,\beta^j;\ \ (i+j)\,\mu_{i+j-1}\,|_{i,j=0}^{r+1} \quad (r = 0, 1, \ldots, [\tfrac{1}{2}(n-1)])$$

are positive.

* These results apply strictly only if the spectrum of the distribution function does not reduce to a finite set of points; in the contrary case the determinants will vanish for large r.

Case (*c*). Terminals $t_1 = -1$, $t_2 = +1$. If $n = 2m$, the determinants

$$\Delta_r(\nu), \quad \Delta_r^*(\nu) = |\beta^j; \quad (i+j-1)\mu_{i+j-2} - (i+j+1)\mu_{i+j}|_{i,j=0}^{r} \quad (r = 0, 1, \ldots, m)$$

are positive; if $n = 2m+1$, the determinants

$$\left.\begin{aligned} \Delta_r^{\text{III}}(\nu) &= |\beta^j; \quad (i+j-1)\mu_{i+j-2} + (i+j)\mu_{i+j-1}|_{i,j=0}^{r+1} \\ \Delta_r^{\text{IV}}(\nu) &= |\beta^j; \quad (i+j-1)\mu_{i+j-2} - (i+j)\mu_{i+j-1}|_{i,j=0}^{r+1} \end{aligned}\right\} \quad (r = 0, 1, \ldots, m)$$

are positive.

6. (i) Thus in case (*c*) (terminals -1, $+1$), where the moments $\mu_0 = 1$, μ_1, μ_2 are known, the first few conditions reduce to

$$(\beta - \mu_1)^2 < 3(\mu_2 - \mu_1^2) < 1 - 3\mu_1^2 + 2\mu_1\beta,$$

and hence we must have (since $-1 \leqslant \beta \leqslant 1$)

$$0 < 3(\mu_2 - \mu_1^2) < \begin{cases} (1-\mu_1)(1+3\mu_1) & (\mu_1 > 0), \\ (1+\mu_1)(1-3\mu_1) & (\mu < 0). \end{cases}$$

These are inequalities for the moments μ_1, μ_2 when the terminals are given as -1, $+1$. We may write them also as inequalities for the terminals t_1, t_2 when the moments are given as $\mu_1 = 0$, $\mu_2 = 1$; we find

$$3 + 2t_1t_2 + t_2^2 < 0 \quad (t_1 + t_2 < 0),$$
$$3 + 2t_1t_2 + t_1^2 < 0 \quad (t_1 + t_2 > 0).$$

(ii) In case (*b*) (lower terminal zero) the first few conditions are:

$$0 < 2\mu_1 - \beta, \quad 0 < 3(\mu_2 - \mu_1^2) - (\beta - \mu_1)^2,$$
$$0 < 8\mu_1\mu_3 - 9\mu_2^2 + 2\beta(3\mu_1\mu_2 - 2\mu_3) + \beta^2(3\mu_2 - 4\mu_1^2),$$
$$0 < 15\mu_2\mu_4 - 16\mu_3^2 + 48\mu_1\mu_2\mu_3 - 20\mu_1^2\mu_4 - 27\mu_2^3$$
$$- \beta(16\mu_1^2\mu_3 + 12\mu_2\mu_3 - 10\mu_1\mu_4 - 18\mu_1\mu_2^2)$$
$$+ \beta^2(8\mu_1\mu_3 + 9\mu_2^2 - 5\mu_4 - 12\mu_1^2\mu_2) - \beta^3(12\mu_1\mu_2 - 4\mu_3 - 8\mu_1^3).$$

7. Suppose y is a random variable taking only the values $0, 1, \ldots$ and having moments (about zero) $\eta_0, \eta_1, \ldots$, and u is an independent rectangular variable (uniform in $(0, 1)$). Then $x = y + u$ has moments

$$\mu_r = \sum_{j=0}^{r} \binom{r}{j} \frac{\eta_j}{j+1} = \frac{1}{r+1} \sum_{j=1}^{r} \binom{r+1}{j} \eta_j,$$

and x is unimodal only if y is. In this case we may take the mode β of x to be at the mode of y.

The inequalities of § 6 (ii) will now give conditions on the moments and mode of y; these conditions will not be the best possible, as in the deriviation of the inequalities of § 6 (ii) we have not used the additional restriction that x must have a 'histogram' distribution. The error will be small. We obtain the following inequalities:

$$0 < 2\eta_1 + 1 - \beta, \quad 0 < 3(\eta_2 - \eta_1^2) + \tfrac{1}{4} - (\beta - \eta_1 - \tfrac{1}{2})^2,$$
$$0 < 8\eta_1\eta_3 - 9\eta_2^2 + 4\eta_3 - 6\eta_1\eta_2 - \eta_1^2 - \beta(4\eta_3 - 6\eta_1\eta_2 + 3\eta_2 - 6\eta_1^2 - \eta_1) + \beta^2(3\eta_2 - 4\eta_1^2 - \eta_1).$$

A general investigation of the implications of these inequalities is complicated by the fact that we cannot standardize the moments; 'location' and 'scale' are determined by the lower terminal (zero) and the interval (unity).

Anscombe compared four distributions having $\eta_1 = 20$, $\eta_2 = 620$ (or 618 in case (iv)), namely: (i) negative binomial, (ii) Polya–Aeppli, (iii) Neyman Type A, (iv) Thomas. Of these, (i) has one mode, (ii) two, and (iii) and (iv) each three. The third and fourth moments can easily be obtained from the generating functions, and substituted in the inequalities above. We find that the moments of even the Thomas distribution could belong to a unimodal distribution, with mode β in the interval $(0, 27)$. Put another way, we cannot deduce *from these moments alone* that the distribution must be multimodal.

Taking more moments will of course give closer limits for β; and if the distribution is in fact multimodal, then for a sufficiently large number of moments the inequalities will not be satisfied for any β.

REFERENCES

ANSCOMBE, F. J. (1950). Sampling theory of the negative binomial and logarithmic series distributions. *Biometrika*, **37**, 358–82.

JOHNSON, N. L. & ROGERS, C. A. (1951). The moment problem for unimodal distributions. *Ann. Math. Statist.* **22**, 433–9.

SHOHAT, J. A. & TAMARKIN, J. D. (1943). The problem of moments. *Mathematical Surveys*, no. 1. American Mathematical Society.

On inverting a class of patterned matrices

BY S. N. ROY AND A. E. SARHAN*

Institute of Statistics, University of North Carolina

1. *Introduction.* We shall use here bold capital letters for matrices and primed letters for their transposes, small letters for scalars, small bold letters for column vectors, small bold primed letters for row vectors. For example, $\mathbf{a}$ will stand for a column vector whose typical element is a_i and $(\mathbf{ab/c})$ for a column vector whose typical element is a_ib_i/c_i. $\mathbf{D}_{a_i}(k\times k)$ will stand for a diagonal matrix whose diagonal elements are $(a_1, a_2, \ldots, a_k)$ and $\tilde{\mathbf{M}}(k\times k)$ will stand for a lower triangular matrix. If all elements of $\mathbf{M}$ are unity we shall write $\mathbf{M}(k\times l)$ as $\mathbf{J}(k\times l)$.

2. If

$$\tilde{\mathbf{M}}_{a_ib_j}(k\times k)=\begin{pmatrix} a_1b_1 & 0 & 0 & \ldots & 0\\ a_2b_1 & a_2b_2 & 0 & \ldots & 0\\ \ldots & \ldots & \ldots & \ldots & \ldots\\ a_kb_1 & a_kb_2 & a_kb_3 & \ldots & a_kb_k \end{pmatrix}, \tag{2·1}$$

then it is easy to check that

$$\tilde{\mathbf{M}}^{-1}_{a_ib_j}(k\times k)=\begin{pmatrix} 1/b_1a_1 & 0 & 0 & \ldots & \ldots & 0 & 0\\ -1/b_2a_1 & 1/b_2a_2 & 0 & \ldots & \ldots & 0 & 0\\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots\\ 0 & 0 & 0 & \ldots & -1/b_{k-1}a_{k-2} & 1/b_{k-1}a_{k-1} & 0\\ 0 & 0 & \ldots & \ldots & \ldots & -1/b_ka_{k-1} & 1/b_ka_k \end{pmatrix}. \tag{2·2}$$

3. If

$$\mathbf{C}=\begin{pmatrix} c_1^2a_1 & c_1c_2a_1 & c_1c_3a_1 & \ldots & c_1c_na_1\\ c_1c_2a_1 & c_2^2(a_1+a_2) & c_2c_3(a_1+a_2) & \ldots & c_2c_n(a_1+a_2)\\ c_1c_3a_1 & c_2c_3(a_1+a_2) & c_3^2(a_1+a_2+a_3) & \ldots & c_3c_n(a_1+a_2+a_3)\\ \ldots & \ldots & \ldots & \ldots & \ldots\\ c_1c_na_1 & c_2c_n(a_1+a_2) & c_3c_n(a_1+a_2+a_3) & \ldots & c_n^2(a_1+a_2+\ldots+a_n) \end{pmatrix}, \tag{3·1}$$

then

$$\mathbf{C}^{-1}=\begin{pmatrix} \frac{1}{c_1^2}\left(\frac{1}{a_1}+\frac{1}{a_2}\right) & -\frac{1}{c_1c_2a_2} & 0 & 0 & \ldots & 0 & 0\\ -\frac{1}{c_1c_2a_2} & \frac{1}{c_2^2}\left(\frac{1}{a_2}+\frac{1}{a_3}\right) & -\frac{1}{c_2c_3a_3} & 0 & \ldots & 0 & 0\\ 0 & -\frac{1}{c_2c_3a_3} & \frac{1}{c_3^2}\left(\frac{1}{a_3}+\frac{1}{a_4}\right) & -\frac{1}{c_3c_4a_4} & \ldots & 0 & 0\\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots\\ 0 & 0 & 0 & 0 & \ldots & -\frac{1}{c_{n-1}c_na_n} & \frac{1}{c_n^2a_n} \end{pmatrix}. \tag{3·2}$$

To prove this, we put $\mathbf{C}=\tilde{\mathbf{M}}_{c_i}(n\times n)\,\mathbf{D}_{a_i}(n\times n)\,\tilde{\mathbf{M}}_{c_i}'(n\times n)$, invert both sides and use (2·1) and (2·2).

* This research was supported, in part, by the United States Air Force through the Office of Scientific Research of the Air Research and Development Command.

Example 1. We need to invert for various purposes the following symmetric positive-definite matrix which is the variance-covariance matrix of order statistics for a sample of size n from an exponential population (Sarhan, 1954):

$$\mathbf{V}(n\times n)=\begin{pmatrix} \frac{1}{n^2} & \frac{1}{n^2} & \frac{1}{n^2} & \dots & \frac{1}{n^2} \\ \frac{1}{n^2} & \frac{1}{n^2}+\frac{1}{(n-1)^2} & \frac{1}{n^2}+\frac{1}{(n-1)^2} & \dots & \frac{1}{n^2}+\frac{1}{(n-1)^2} \\ \frac{1}{n^2} & \frac{1}{n^2}+\frac{1}{(n-1)^2} & \sum_{i=1}^{3}\frac{1}{(n-i+1)^2} & \dots & \sum_{i=1}^{3}\frac{1}{(n-i+1)^2} \\ \dots & \dots & \dots & \dots & \dots \\ \frac{1}{n^2} & \frac{1}{n^2}+\frac{1}{(n-1)^2} & \sum_{i=1}^{3}\frac{1}{(n-i+1)^2} & \dots & \sum_{i=1}^{n}\frac{1}{(n-i+1)^2} \end{pmatrix}. \tag{3·3}$$

This can be put in the form (3·1) by putting

$$c_1=c_2=\dots=c_n=1$$

and

$$a_1=\frac{1}{n_1},\quad a_2=\frac{1}{(n-1)^2},\quad a_3=\frac{1}{(n-2)^2},\quad \dots,\quad a_n=1. \tag{3·4}$$

Therefore, by substituting in (3·2), we have

$$\mathbf{V}^{-1}=\begin{pmatrix} n^2+(n-1)^2 & -(n-1)^2 & 0 & 0 & \dots & 0 & 0 \\ -(n-1)^2 & (n-1)^2+(n-2)^2 & -(n-2)^2 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \dots & -1 & 1 \end{pmatrix}. \tag{3·5}$$

This inverse can also be obtained by induction.

4. If

$$\mathbf{C}(k\times k)=[\mathbf{D}_{p_i}(k\times k)+\lambda\mathbf{q}(k\times 1)\,\mathbf{r}'(1\times k)], \tag{4·1}$$

then

$$\mathbf{C}^{-1}=\mathbf{D}_{1/p_i}-\frac{\lambda}{\left(1+\lambda\sum_{i=1}^{k}q_ir_i/p_i\right)}(\mathbf{q}/\mathbf{p})\,(\mathbf{r}/\mathbf{p})'. \tag{4·2}$$

Proof. Assume that

$$\mathbf{C}^{-1}=\mathbf{D}_{1/p_i}(k\times k)+\mu(\mathbf{q}/\mathbf{p})\,(\mathbf{r}/\mathbf{p})', \tag{4·3}$$

where μ is unknown. Therefore

$$\begin{aligned}\mathbf{I}(k\times k)=\mathbf{C}\mathbf{C}^{-1}&=(\mathbf{D}_p+\lambda\mathbf{q}\mathbf{r}')\,(\mathbf{D}_{1/p_i}+\mu(\mathbf{q}/\mathbf{p})\,(\mathbf{r}/\mathbf{p})')\\ &=\mathbf{I}+\mu\mathbf{q}(\mathbf{r}/\mathbf{p})'+\lambda\mathbf{q}(\mathbf{r}/\mathbf{p})'+\lambda\mu\sum_{i=1}^{k}(q_ir_i/p_i)\,\mathbf{q}(\mathbf{r}/\mathbf{p})'\\ &=\mathbf{I}+\left(\mu+\lambda+\lambda\mu\sum_{i=1}^{k}q_ir_i/p_i\right)\mathbf{q}(\mathbf{r}/\mathbf{p})'.\end{aligned}$$

Thus

$$\mu\left(1+\lambda\sum_{i=1}^{k}q_ir_i/p_i\right)+\lambda=0 \tag{4·4}$$

or

$$\mu=-\lambda\Big/\left(1+\lambda\sum_{i=1}^{k}q_ir_i/p_i\right),$$

which completes the proof of (4·2).

Example 2. We need to invert the following symmetric positive-definite matrix which is the variance-covariance matrix of the multinomial distribution:

$$\mathbf{V}(k\times k)=\begin{pmatrix} p_1(1-p_1) & -p_1p_2 & -p_1p_3 & \dots & -p_1p_k \\ -p_1p_2 & p_2(1-p_2) & -p_2p_3 & \dots & -p_2p_k \\ -p_1p_3 & -p_2p_3 & p_3(1-p_3) & \dots & -p_3p_k \\ \dots & \dots & \dots & \dots & \dots \\ -p_1p_k & -p_2p_k & -p_3p_k & \dots & -p_k(1-p_k) \end{pmatrix}. \tag{4·5·1}$$

This can be put in the form

$$\mathbf{D}_{p_i}(k\times k)+\lambda\mathbf{p}(k\times 1)\,\mathbf{p}'(1\times k),$$

where $\lambda=-1$, that is,

$$\mathbf{V}=\begin{pmatrix} p_1 & & & 0\\ & p_2 & & \\ & & \ddots & \\ 0 & & & p_k\end{pmatrix}-\begin{pmatrix} p_1\\ \vdots\\ p_k\end{pmatrix}(p_1\cdots p_k).$$

Therefore, from (4·2) we have

$$\mathbf{V}^{-1}=\mathbf{D}_{1/p_i}+\frac{1}{1-\sum_{i=1}^{k}p_i}\mathbf{J}$$

$$=\begin{pmatrix} \frac{1}{p_i}+\frac{1}{1-\sum_{i=1}^{k}p_i} & \frac{1}{1-\sum_{i=1}^{k}p_i} & \cdots & \frac{1}{1-\sum_{i=1}^{k}p_i}\\ \frac{1}{1-\sum_{i=1}^{k}p_i} & \frac{1}{p_2}+\frac{1}{1-\sum_{i=1}^{k}p_i} & \cdots & \frac{1}{1-\sum_{i=1}^{k}p_i}\\ \cdots & \cdots & \cdots & \cdots\\ \frac{1}{1-\sum_{i=1}^{k}p_i} & \frac{1}{1-\sum_{i=1}^{k}p_i} & \cdots & \frac{1}{p_k}+\frac{1}{1-\sum_{i=1}^{k}p_i}\end{pmatrix}. \qquad (4\cdot5\cdot2)$$

5. If

$$\mathbf{C}\{(n+p)\times(n+p)\}=\begin{pmatrix}\mathbf{D}_k & a\mathbf{J}\\ a\mathbf{J} & \mathbf{D}_m\end{pmatrix}\begin{matrix}n\\ p\end{matrix}, \qquad (5\cdot1)$$

$$\qquad\qquad\quad n\qquad p$$

then

$$\mathbf{C}^{-1}=\begin{pmatrix}\frac{1}{k}\mathbf{I}+\alpha\mathbf{J} & \beta\mathbf{J}\\ \beta\mathbf{J} & \frac{1}{m}\mathbf{I}+\gamma\mathbf{J}\end{pmatrix}\begin{matrix}n\\ p\end{matrix}, \qquad (5\cdot2)$$

$$\qquad\qquad\quad n\qquad p$$

where $$\alpha=-\frac{a^2p}{k(na^2p-mk)},\quad \beta=\frac{a}{na^2p-mk}\quad\text{and}\quad \gamma=-\frac{a^2n}{m(na^2p-mk)}. \qquad (5\cdot3)$$

Proof. Assuming $\mathbf{C}^{-1}$ to be of the general structure (5·2) and equating $\mathbf{C}\mathbf{C}^{-1}$ to $\mathbf{I}\{(n+p)\times(n+p)\}$ we have

$$\mathbf{I}=\begin{pmatrix}\frac{1}{k}\mathbf{D}_k+\alpha k\mathbf{J}+pa\beta\mathbf{J} & \beta k\mathbf{J}+\frac{a}{m}\mathbf{J}+a\gamma p\mathbf{J}\\ \beta k\mathbf{J}+\frac{a}{m}\mathbf{J}+a\gamma p\mathbf{J} & \frac{1}{m}\mathbf{D}_m+\gamma m\mathbf{J}+a\beta n\mathbf{J}\end{pmatrix}\begin{matrix}n\\ p\end{matrix}$$

$$\qquad\qquad\qquad\quad n\qquad\qquad\quad p$$

or $$\alpha k+pa\beta=0,\quad \beta k+\frac{a}{m}+a\gamma p=0,\quad \gamma m+an\beta=0, \qquad (5\cdot4)$$

from which, by solving for α, β and γ, we have (5·3) and thus (5·2).

Example 3. We need to invert the following non-singular matrix which occurs in analysis of variance for two-way classification with equal frequencies in the different cells

$$\mathbf{C}=\begin{pmatrix}\mathbf{D}_k & \mathbf{J}\\ \mathbf{J} & \mathbf{D}_m\end{pmatrix}\begin{matrix}m\\ k-1\end{matrix}. \qquad (5\cdot5)$$

$$\qquad\quad m\qquad k-1$$

This is easily seen to be a special case of (5·1) by putting $a = 1$, $n = m$, $p = k-1$. Now substituting in (5·3) and (5·2) we have

$$\mathbf{C}^{-1} = \begin{pmatrix} \frac{1}{k}\mathbf{I} + \frac{k-1}{mk}\mathbf{J} & -\frac{1}{m}\mathbf{J} \\ -\frac{1}{m}\mathbf{J} & \frac{1}{m}\mathbf{I} + \frac{1}{m}\mathbf{J} \end{pmatrix} \begin{matrix} m \\ k-1 \end{matrix} \qquad (5\cdot6)$$

$$\begin{matrix} m & k-1 \end{matrix}$$

6. If

$$\mathbf{C}\{(3k-2)\times(3k-2)\} = \begin{pmatrix} \mathbf{D}_k & \mathbf{J} & \mathbf{J} \\ \mathbf{J} & \mathbf{D}_k & \mathbf{J} \\ \mathbf{J} & \mathbf{J} & \mathbf{D}_k \end{pmatrix} \begin{matrix} k \\ k-1, \\ k-1 \end{matrix} \qquad (6\cdot1)$$

$$\begin{matrix} k & k-1 & k-1 \end{matrix}$$

then

$$\mathbf{C}^{-1} = \begin{pmatrix} \mathbf{D}_{1/k} + \alpha\mathbf{J} & \beta\mathbf{J} & \gamma\mathbf{J} \\ \beta\mathbf{J} & \mathbf{D}_{1/k} + \delta\mathbf{J} & \phi\mathbf{J} \\ \gamma\mathbf{J} & \phi\mathbf{J} & \mathbf{D}_{1/k} + \xi\mathbf{J} \end{pmatrix} \begin{matrix} k \\ k-1, \\ k-1 \end{matrix} \qquad (6\cdot2)$$

$$\begin{matrix} k & k-1 & k-1 \end{matrix}$$

where $$\alpha = 2(k-1)/k^2, \quad \beta = \gamma = -1/k, \quad \delta = \xi = 1/k \quad \text{and} \quad \phi = 0. \qquad (6\cdot3)$$

Proof. Assuming $\mathbf{C}^{-1}$ to be of the general structure (6·2) and equating $\mathbf{C}\mathbf{C}^{-1}$ to $I\{(3k-2)\times(3k-2)\}$, we have, as in the case of (6·2), six linear equations in the six unknowns $\alpha, \beta, \gamma, \delta, \phi, \xi$, from which, by solving, we have the unknowns in the form (6·3).

The matrices $\mathbf{C}$ and $\mathbf{C}^{-1}$ of (6·1) and (6·2) occur in analysis of variance for Latin squares.

7. If

$$\mathbf{C}(k\times k) = \begin{pmatrix} a & b & b & b & \dots & b \\ b & c & d & d & \dots & d \\ b & d & c & d & \dots & d \\ \dots & \dots & \dots & \dots & \dots & \dots \\ b & d & d & \dots & \dots & c \end{pmatrix}, \qquad (7\cdot1)$$

then

$$\mathbf{C}^{-1} = \begin{pmatrix} e & f & f & f & \dots & f \\ f & g & h & h & \dots & h \\ f & h & g & h & \dots & h \\ \dots & \dots & \dots & \dots & \dots & \dots \\ f & h & h & h & \dots & g \end{pmatrix}, \qquad (7\cdot2)$$

where $$e = \frac{1}{a}[1+\lambda b^2(k-1)], \quad f = -\lambda b, \quad g = \frac{1}{c-d}[1-\lambda(ad-b^2)], \quad h = \lambda(b^2-ad)/(c-d), \qquad (7\cdot3)$$

and $$\lambda = 1/[a(c-d)+(k-1)(ad-b^2)]. \qquad (7\cdot4)$$

Proof. The right-hand side of (7·1) can be written in the form

$$\begin{pmatrix} a & b\mathbf{1}' \\ b\mathbf{1} & \mathbf{D}_{c-d} + d\mathbf{J} \end{pmatrix} \begin{matrix} 1 \\ k-1 \end{matrix}, \qquad (7\cdot5)$$

$$\begin{matrix} 1 & k-1 \end{matrix}$$

We now assume that the right-hand side of (7·2) can be written in the form

$$\begin{pmatrix} e & f\mathbf{1}' \\ f\mathbf{1} & \mathbf{D}_{g-h} + h\mathbf{J} \end{pmatrix} \begin{matrix} 1 \\ k-1 \end{matrix}, \qquad (7\cdot6)$$

$$\begin{matrix} 1 & k-1 \end{matrix}$$

where e, f, g, h are undetermined, and $\mathbf{1}' = (1, 1, \dots, 1)$.

Multiplying (7·5) with (7·6) and equating to $\mathbf{I}(k\times k)$, we have

$$\begin{pmatrix} 1 & 0 \\ 0 & \mathbf{I} \end{pmatrix} \begin{matrix} 1 \\ k-1 \end{matrix} = \begin{pmatrix} ae + bf(k-1) & af\mathbf{1}' + b(g-h)\mathbf{1}' + bh(k-1)\mathbf{1}' \\ af\mathbf{1} + b(g-h)\mathbf{1} + bh(k-1)\mathbf{1} & bf\mathbf{J} + (c-d)(g-h)\mathbf{I} + h(c-k)\mathbf{J} \\ & \quad + d(g-h)\mathbf{J} + dh(k-1)\mathbf{J} \end{pmatrix}. \qquad (7\cdot7)$$

$$\begin{matrix} 1 & k-1 \end{matrix}$$

Thus we should have

$$ae+bf(k-1)=1, \quad af+b(g-h)+bh(k-1)=0, \quad (c-d)(g-h)=1,$$

$$bf+h(c-d)+d(g-h)+dh(k-1)=0, \tag{7·8}$$

solving which we obtain (7·3).

The matrix (7·1) is common in response surfaces. 5, 6 and 7 can also be subsumed under 4.

The authors would like to thank Prof. M. G. Kendall for his valuable editorial suggestions.

REFERENCE

SARHAN, A. E. (1954). Estimation of the mean and the standard deviation by order statistics. *Ann. Math. Statist.* **25**, 317.

A note on the risks of error involved in the sequential ratio test

BY J. MEDHI

Institut de Statistique, University of Paris

1. INTRODUCTION

For testing hypotheses, the sequential procedure, formulated by Wald, at first involves the successive calculation of

$$Z_1=z_1, \quad Z_2=z_1+z_2, \ldots Z_n=z_1+\ldots+z_n,$$

where

$$z_i=z(x_i)=\log\frac{f_0(x_i)}{f_1(x_i)},$$

the densities $f_0(x), f_1(x)$ (the probabilities p_0, p_1 in case of discrete distribution) of x corresponding to the two alternative hypotheses H_0, H_1 respectively. Then, if at any stage $Z \geqslant a$, H_0 is accepted; if $Z \leqslant b$, H_1 is accepted and in case $b<Z<a$, a new independent sample* is to be taken. Let the chance of error in accepting H_1 when H_0 is true be α and that in accepting H_0 when H_1 is true be β. Then we have approximately $a=\log(1-\alpha)-\log\beta$ and $b=\log\alpha-\log(1-\beta)$.

That the sequential analysis theory is also applicable to a single sample was first pointed out by G. A. Barnard. As a solution to the important problem of discrimination between two types of spectra—continuous and discrete—of time series, a test procedure based on such an application has been suggested by Bartlett (1954). Here the criterion is the log likelihood ratio, $\log(p_0/p_1)$, where the probabilities p_0, p_1 for the sample for each of the hypotheses, H_0 (discrete spectrum), H_1 (continuous spectrum) have to be maximized with respect to the unknown parameters involved in the specifications. We regard the entire available sample of n observations as the first of a hypothetical sequence of independent samples. Then if we take the risks of error in accepting H_1 when H_0 is true and vice versa to be equal, being at most not greater than ϵ, we may, following sequential theory, adopt the decision rule:

(i) if $\log(p_0/p_1) \geqslant \log\dfrac{1-\epsilon}{\epsilon}$, then accept H_0,

(ii) if $\log(p_0/p_1) \leqslant \log\dfrac{\epsilon}{1-\epsilon}$, then accept H_1,

(iii) if $\log\dfrac{1-\epsilon}{\epsilon} > \log(p_0/p_1) > \log\dfrac{\epsilon}{1-\epsilon}$, then reach no decision (see also Medhi (1956)).

Now when the probability of a decision with the single sample available is not small, such a procedure might lead to a rather over-cautious assessment of ma ·num risks ϵ. It is therefore advisable to investigate the more precise risks of error in such circumstances; in ordinary sequential analysis we know that the usual procedure gives a good approximation in most cases. In this note an explicit example has been considered to obtain such risks of error in the circumstances stated.

* The symbol x may denote a *group* of observations, i.e. a 'sample'.

2. Mean of a normal sample

Let
$$f(x) = \frac{1}{\sqrt{(2\pi\sigma^2)}} \exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\} \quad (m>0), \tag{2·1}$$

and let the two alternative hypotheses be

$$H_0\text{: mean} = +m, \qquad H_1\text{: mean} = -m.$$

Then
$$z_i = z(x_i) = \log\frac{f_0(x_i)}{f_1(x_i)} = \frac{2m}{\sigma^2}x_i. \tag{2·2}$$

We suppose that the hypothesis H_0 obtains.

Let $p_\gamma^{(i)}(p_\omega^{(i)})$ denote the probability of arriving at the right (wrong) decision at the ith step,

$$p_\gamma^{[m]} \equiv p_\gamma^{(1)} + \ldots + p_\gamma^{(m)}, \quad p_\omega^{[m]} \equiv p_\omega^{(1)} + \ldots + p_\omega^{(m)}.$$

Now $g(z)$, the frequency function of z, is given by

$$g(z) = \frac{1}{\sqrt{(2\pi\sigma_1^2)}} \exp\left\{-\frac{(z-\frac{1}{2}\sigma_1^2)^2}{2\sigma_1^2}\right\} \quad (\sigma_1 = 2m/\sigma). \tag{2·3}$$

The frequency function $g_1(Z_1)$ of $Z_1 = z_1$ is of exactly the same form as $g(z)$, so that we have

$$p_\gamma^{(1)} = \int_a^\infty g_1(Z_1)\,dZ_1 = \int_a^\infty g(z)\,dz, \tag{2·4}$$

and
$$p_\omega^{(1)} = \int_{-\infty}^b g_1(Z_1)\,dZ_1 = \int_{-\infty}^b g(z)\,dz. \tag{2·5}$$

$g_n(Z_n)$, the conditional distribution of $Z_n = z_1 + \ldots + z_n$, is given for $n>1$ by the 'truncated convolution' (Samuelson, 1948),

$$g_n(Z_n) = \int_b^a g(Z_n - s)\, g_{n-1}(s)\,ds. \tag{2·6}$$

We have here

$$\begin{aligned} g_2(Z_2) &= \int_b^a g(Z_2 - s)\, g_1(s)\,ds \quad (-\infty < Z_2 < \infty) \\ &= \frac{1}{\sqrt{(2\pi\sigma_1^2)}} \exp\left\{-\frac{\frac{1}{2}(Z_2-\sigma_1^2)^2}{2\sigma_1^2}\right\} \cdot \frac{1}{\sqrt{2}}\left[F\left\{\frac{\sqrt{2}}{\sigma_1}(\tfrac{1}{2}Z_2 - b)\right\} - F\left\{\frac{\sqrt{2}}{\sigma_1}(\tfrac{1}{2}Z_2 - a)\right\}\right], \end{aligned} \tag{2·7}$$

where
$$F(x) = \int_{-\infty}^x f(t)\,dt.$$

Hence
$$p_\gamma^{(2)} = \int_a^\infty g_2(Z_2)\,dZ_2 \tag{2·8}$$

and
$$p_\omega^{(2)} = \int_{-\infty}^b g_2(Z_2)\,dZ_2 = \int_{-\infty}^{b_2} \left\{\frac{1}{\sqrt{(2\pi)}} \exp(-\tfrac{1}{2}t^2)\right\}\{F(t+c') - F(t+c)\}\,dt, \tag{2·9}$$

where
$$b_2 = (b-\sigma_1^2)/(\sqrt{2}\sigma_1), \quad c = (\sigma_1^2 - 2a)/(\sqrt{2}\sigma_1), \quad c' = (\sigma_1^2 - 2b)/(\sqrt{2}\sigma_1).$$

Again
$$\begin{aligned} g_3(Z_3) &= \int_b^a g(Z_3 - s)\, g_2(s)\,ds \\ &= \frac{1}{(2\sqrt{2\pi\sigma_1^2})} \int_b^a \exp\left\{-\frac{\frac{1}{3}(Z_3 - \frac{3}{2}\sigma_1^2)^2}{2\sigma_1^2}\right\} \\ &\quad \times \exp\left\{-\frac{\frac{3}{2}(s - \frac{2}{3}Z_3)^2}{2\sigma_1^2}\right\} \times \left[F\left\{\frac{\sqrt{2}}{\sigma_1}(a - \tfrac{1}{2}s)\right\} - F\left\{\frac{\sqrt{2}}{\sigma_1}(b - \tfrac{1}{2}s)\right\}\right] ds, \end{aligned} \tag{2·10}$$

so that
$$p_\gamma^{(3)} = \int_a^\infty g_3(Z_3)\,dZ_3, \tag{2·11}$$

and
$$p_\omega^{(3)} = \int_{-\infty}^b g_3(Z_3)\,dZ_3 = \int_{-\infty}^{b_3}\int_b^a \frac{\sqrt{3}}{\sqrt{2}\sigma_1} K_1 K_2 K_3\,ds\,dt, \tag{2·12}$$

where

$$K_1 = \frac{1}{\sqrt{(2\pi)}}\exp(-\tfrac{1}{2}t^2), \quad K_2 = \frac{1}{\sqrt{(2\pi)}}\exp\left\{-\frac{\frac{3}{2}\left(s-\sigma_1^2-\frac{2\sqrt{3}}{3}\sigma_1 t\right)^2}{2\sigma_1^2}\right\},$$

$$K_3 = F\{\sqrt{2}(a-\tfrac{1}{2}s)/\sigma_1\} - F\{\sqrt{2}(b-\tfrac{1}{2}s)/\sigma_1\}, \quad b_2 = \sqrt{3}\,b/(3\sigma_1) - \tfrac{1}{2}\sqrt{3}\,\sigma_1.$$

Proceeding precisely in the same way, we can calculate $g_4(Z_4)$ and then $p_\gamma^{(4)}$, $p_\omega^{(4)}$ and so on.

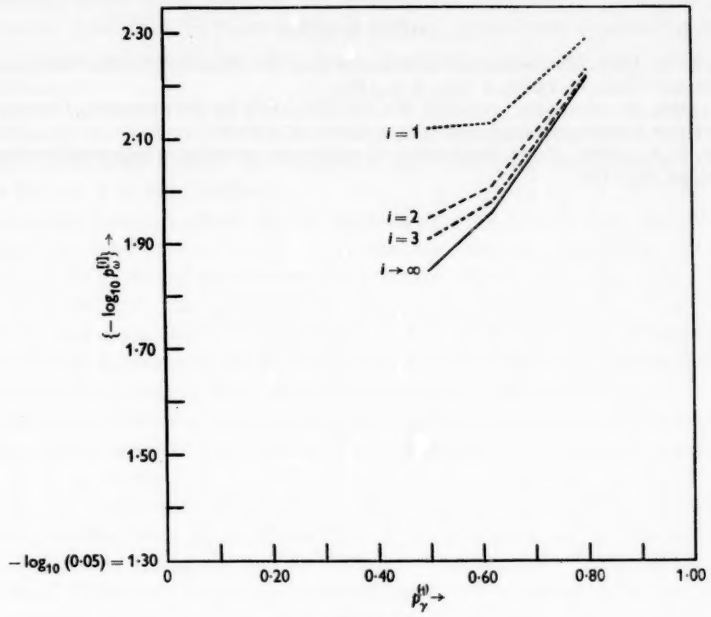


Fig. 1

3. Numerical illustration

We take $\epsilon = 0{\cdot}05$ and choose σ_1 such that $p_\gamma^{(1)}$ is fairly high.

(i) Let $p_\gamma^{(1)} = 0{\cdot}7995$, then $\sigma_1 = 3{\cdot}4080$.

Calculations give

$$p_\omega^{(1)} = 0{\cdot}00511, \quad p_\omega^{(2)} = 0{\cdot}00077, \quad p_\omega^{(3)} = 0{\cdot}00010$$

so that

$$p_\omega^{[3]} = 0{\cdot}00598.$$

A rough value of $p_\omega^{[\infty]}$ obtained graphically is 0·006.

(ii) Let $p_\gamma^{(1)} = 0{\cdot}6206$, then $\sigma_1 = 2{\cdot}7006$.

We have

$$p_\omega^{(1)} = 0{\cdot}00733, \quad p_\omega^{(2)} = 0{\cdot}00246, \quad p_\omega^{(3)} = 0{\cdot}00063,$$

and thus $p_\omega^{[3]} = 0{\cdot}01042$ and graphical evaluation gives $p_\omega^{[\infty]} = 0{\cdot}011$ roughly.

(iii) Let σ_1 be such that $p_\omega^{(1)}$ is maximum, which implies that $p_\gamma^{(1)} = 0{\cdot}50$ and $\sigma_1 = \sqrt{(2a)} = 2{\cdot}4267$. In this case

$$p_\omega^{(1)} = 0{\cdot}0075, \quad p_\omega^{(2)} = 0{\cdot}0036, \quad p_\omega^{(3)} = 0{\cdot}0011,$$

giving $p_\omega^{[3]} = 0{\cdot}0122$, and $p_\omega^{[\infty]}$ obtained graphically is 0·014 roughly.

In each of the three cases, $p_\omega^{(2)}$ and $p_\omega^{(3)}$ were evaluated by numerical methods. We considered the intervals (d_2, b_2) instead of $(-\infty, b_2)$ in case of $p_\omega^{(2)}$ and (d_3, b_3) instead of $(-\infty, b_3)$ in the case of $p_\omega^{(3)}$; d_2, d_3 were chosen such that the corresponding integrands become negligibly small for values of the argument less than d_2 in the former case and for values less than d_3 in the latter case.

We find that when the probability of the right decision at the first step is 0·7995, 0·6206, 0·50, the total risks of error amount to roughly 0·006, 0·011 and 0·014 respectively, as against the maximum risk $\epsilon = 0{\cdot}05$.

The numerical results have been graphically illustrated by drawing the graphs $-\log p_\omega^{(i)}$ against $p_\gamma^{(1)}$ as abscissa (for $i = 1, 2, 3, \infty$).

My sincere thanks are due to Prof. M. S. Bartlett for suggesting to me this problem.

REFERENCES

BARTLETT, M. S. (1954). Problèmes de l'analyse spectrale des séries temporelles stationnaires. *Publ. Inst. Statist.* (Univ. de Paris), **3**, fasc. 3, p. 119.

MEDHI, J. (1956). A note on the properties of a test procedure for discrimination between two types of spectra of a stationary stochastic process (to be published).

SAMUELSON, P. A. (1948). Exact distribution of continuous variables in sequential analysis. *Econometrica*, **16**, 2, p. 191.

CORRIGENDA

(1) GUNNAR BLOM. 'Transformations of the binomial, negative binomial, Poisson and χ^2 distributions', *Biometrika* (1954), **41**, 302.

I am indebted to Mr A. R. Thatcher for drawing my attention to a computational slip which has led to three errors in Table 1 on p. 310 of my paper which appeared in *Biometrika* (1954). In the last column of the table, lower confidence limits are given, being roots of the equation in p

$$a - \tfrac{1}{2} - np = \lambda_\varepsilon \sqrt{(npq)}. \qquad (7{\cdot}5)$$

The entries for $a = 5$ should be 0·051, 0·038 and 0·030 instead of 0·002, 0·001 and < 0 respectively. As a consequence, the remark on the same page that (7·5) is 'remarkably inaccurate for $a = 5$' is not justified.

The correction does not affect the recommendation on p. 310 that the inverse sine formula should be used instead of (7·5). It is seen from the corrected Table 1 and has also been confirmed by extended calculations that the inverse sine formula, even without the skewness correction given in formula (7·3), provides more accurate values for confidence levels of 90 % and more. For low confidence levels (7·5) is better, the difference between the formulae being, however, small. These facts, together with the computational simplicity of the inverse sine formula, lead to the recommendation mentioned above.

In this connexion, it should be mentioned that the extended calculations have proved the tentative statement on p. 310 that formula (7·4) is preferable to (7·5) to be incorrect for small sample sizes. If, for example, $n = 20$ (7·5) is, in fact, more accurate. In view of the advantages of the inverse sine formula referred to above, this comparison is, however, of little interest.

It might be added that C. A. Bennett & N. L. Franklin, on p. 606 of their book *Statistical Analysis in Chemistry and the Chemical Industry* (New York: Wiley; London: Chapman Hall, 1954) give a quadratic expression due to Freeman & Tukey, which also provides confidence limits for a binomial probability. The roots of the quadratic can, with the notation used in my paper, be written

$$p = p^* \pm c_n \lambda_\varepsilon \sqrt{\left(\frac{p^* q^*}{n+1}\right) + \frac{\lambda_\varepsilon^2}{4(n+1)}} (q^* - p^*),$$

where

$$c_n = \sqrt{\left(1 - \frac{\lambda_\varepsilon^2}{4(n+1)}\right)}.$$

This formula bears a certain resemblance to the Cornish–Fisher expansion (7·2) on p. 309 of my paper, but it is much more accurate than (7·2) for small sample sizes, even when the skewness correction of the last-mentioned formula is used. The performance of the Bennett & Franklin formula is about the same as that of the inverse sine formula (without skewness correction), the last formula being, however, in general slightly better. Therefore, it seems unwarranted to use the method described by Bennett & Franklin, except when it is considered to be an advantage to avoid using a table of the inverse sine function.

(2) M. B. WILK, *Biometrika* (1955), **42**, 70.

P. 74, Table 2, second line:

$$\textit{for} \quad \frac{rp}{(t-1)} \sum_{ik} (bt)_{ik}^2 \quad \textit{read} \quad \frac{rp}{(t-1)} \sum_k t_k^2.$$

REVIEWS

Choice and Chance by Cardpack and Chessboard. Volume II. By LANCELOT HOGBEN. London: Max Parrish and Co. Ltd. 1955. Pp. 466. 70*s*.

This is the continuation of the first volume which appeared in 1950. The sub-title to this, as to the first, is 'An Introduction to Probability by Visual Aids', but this, in the reviewer's opinion, is not an accurate description. What Prof. Hogben has done is to write two text-books about statistical methods, illustrating with diagrams at intervals the various facets of the finite fundamental sets from which probability is calculated. As the statistical methods become more complicated the diagrams become more infrequent, and in this second volume Prof. Hogben has to come down to algebra like everyone else, and begins with a chapter on expectation techniques. This is followed by bivariate models, simple analysis of variance, moments, sampling distributions, significance tests, regression, covariance and factor analysis, and an awkward chapter on sampling in a finite universe. The book is completed by a chapter entitled 'Second Thoughts on Significance', which reveals a certain naïvety of outlook on the part of the author.

It is certain that these two books will be of little use to the research worker in other fields who wishes to learn a little statistical method. In spite of the attractively produced diagrams the mathematical approach is there and must be mastered, and this mathematical approach is not rendered less formidable than that of any other statistics text-book by the author's delineation of it. Serious students of mathematical statistics will undoubtedly prefer something more definite, such as, for example, Prof. Kendall's two volumes. They will find in Prof. Hogben's two volumes, however, a certain freshness of exposition which will possibly help them when they come to revise what they already know.

F. N. DAVID

Decision Processes. Edited by R. M. THRALL, C. H. COOMBS and R. L. DAVIS. New York: John Wiley and Sons, Inc.; London: Chapman and Hall Ltd. 1954. Pp. viii + 332. 40*s*.

This book contains edited versions of some of the papers delivered at a seminar on the 'Design of Experiments in Decision Processes' held in the summer of 1952 in Santa Monica, California. The authors 'regard it as a vehicle for raising a number of basic questions and perhaps also providing some guideposts towards answers to some of these questions'.

Perhaps the most important thing for a reviewer in this journal to do is to warn readers that the book is concerned only tangentially with the theory of statistical decision functions in the sense of Wald, and in fact, because of the relative weight of psychological and economic interest in most of the papers, few of them would qualify, by their subject-matter, for publication in *Biometrika*. This is not to say, however, that readers of *Biometrika*, whose broader interests lie in the directions indicated, will not find the papers worth reading. The terse style and self-contained nature of most of the chapters make them well suited to reading at odd moments.

After two introductory chapters, the second devoted to the theory of scales of measurement, there follow four parts devoted, respectively, to 'Individual and Social Choice', 'Learning Theory', 'Theory and Applications of Utility' and 'Experimental Studies'. One of the most interesting papers in the first section discusses the validity of four possible principles for choosing between alternative courses of action, in the light of incomplete information:

(1) Wald's minimax rule, or Manicheeism, which amounts to the belief that this world is the worst of all possible.

(2) A rule suggested by Hurwicz, which is a weighted combination of Manicheeism with the Leibnizian notion that this world is the best of all possible.

(3) Savage's minimax regret rule, in which we assume, not that the world is the worst of all possible, but that the consequences of our ignorance are the worst of all possible.

(4) Laplace's rule, in which we attach the same *a priori* probability to each possible world.

The author of this paper shows that Laplace's rule is perhaps the least unsatisfactory of these four rules, and goes on to consider further possible principles in accordance with certain conditions which he lays down.

A good many of the papers discuss in a highly critical spirit the notion of expected utility in connexion with decisions, and after reading these papers anyone who believed that rational decisions were those which maximized expected utility, in any other than a purely tautological sense, would have his belief severely shaken. The section devoted to Learning Theory carries on the development of this most interesting application of the theory of Conditional Probabilities which is, in the minds of most statisticians, associated with the name of Frederick Mosteller. The experimental studies represent perhaps the weakest part of the book, in that the problems considered are extremely trivial. It is remarkable that no use seems to be made of the very rich sources of experimental material of this kind which exist in the records of human activity which we call history, in a broad sense. But this may be being unfair to the authors, because it seems that most, if not all, the participants in this seminar recognize that their mathematical theories of decision processes must as yet be regarded as very remote from application to real human behaviour.

G. A. BARNARD

Stochastic Models for Learning. By R. R. BUSH and F. MOSTELLER. New York: John Wiley and Sons, Inc.; London: Chapman and Hall Ltd. 1955. Pp. xvi + 365. 72*s*.

With the usual type of learning experiment a subject has to make one of a number of alternative responses at each of a sequence of trials. Some of these responses are rewarded while others are punished, and learning is said to occur if the rewarded responses are made with increasing frequency as the trials proceed. In this book, Bush and Mosteller have attempted to describe such learning behaviour by means of probabiiity models. A suitable model would appear to require that the probability of a particular response occurring at a given trial is some function of all the preceding responses; further, that the mean probability for the occurrence of the rewarded responses should increase, while that of the punished responses decrease, with successive trials. In order to describe what is essentially a non-stationary discrete time series, the authors have constructed their model as a kind of Markov process, where the state of the system at each trial is defined in terms of a vector, giving the probabilities for the occurrence of each type of response. Theories of learning are not sufficiently explicit to dictate the form which such a process should take, and consequently the authors have merely sought to describe their data in the most economical manner. This is in constrast with the physicist and the geneticist, who are in the enviable position of being able to construct stochastic processes which are intimately related to extensive and precisely formulated theories.

The book is divided into two main parts. In the first, a general stochastic process is developed in terms of a system of linear operators, together with the derivation of a number of its mathematical properties. In particular, recurrence formulae are given for the moments of the various response probabilities at the nth trial. The second part deals with the application of special cases of this general process to data obtained from a number of learning experiments performed with animal and human subjects. It is in this latter half of the book that the authors deal with the problems of estimating the parameter values and of assessing the goodness of fit of their models. These problems appear to be of great difficulty and the various attempts which Bush and Mosteller have made to overcome them, while admittedly of an *ad hoc* nature, are of value as possibly encouraging further work in a relatively unexplored field of statistical theory.

A. R. JONCKHEERE

Introduction to Demography. By MORTIMER SPIEGELMAN. Chicago, Illinois: Society of Actuaries. 1955. Pp. xxi + 309. $6.00.

This book was commissioned by the Society of Actuaries for the use of students preparing themselves for the Society's examinations. It is thus a parallel work to that of P. R. Cox (*Demography*, Cambridge University Press, 1950). While Cox is mainly concerned with the collection and analysis of British vital statistics, Spiegelman naturally places more emphasis on data arising in the United States and Canada. There is, for example, a more detailed and comprehensive study of errors occurring in censuses, and other vital statistics, than is given by Cox, since such errors have persisted longer, and to a greater degree, in the American censuses than in those of Great Britain.

The book should be admirably suited to the needs of actuarial students in that the items of necessary information are set out clearly in (usually) short sections. On the other hand, there tends to be the typical text-book drawback of a lack of deep critical discussion.

Subsequent to a discussion of demographic data and their shortcomings there is a chapter on measures of mortality, and comparison of mortality in different populations. The author's comparison

of 'direct' and 'indirect' adjusted rates seems to be less than fair to the latter. A conventional treatment of life-table construction and mortality projections rounds off the treatment of mortality.

There follows a brief chapter, too brief in the opinion of the reviewer, on morbidity statistics and a treatment of family composition and the various indices of fertility and reproduction. The next chapter deals with the distribution of the population, and internal and external migration. Internal migration is considered rather cursorily and recent British work on the balance of internal migration is not mentioned. The following chapter, on 'The Working Population', gives an interesting and clear analysis of the structure of the working population of the United States, based on published material. The final chapter, on population estimates and projections, gives a fair statement of the problems and methods of projection, without, perhaps, sufficient cautionary advice on the value of these projections when calculated.

To sum up, this is a good text-book of real value to students, though it breaks little significantly fresh ground in the demographic field.

N. L. JOHNSON

Mortality and Other Investigations. Volume I. By H. W. HAYCOCKS and W. PERKS. Cambridge University Press. 1955. Pp. ix + 164. 20*s*.

This book is one of a series written for the post-war syllabus of the Institute of Actuaries and forms the first introduction to the more practical subjects included in the actuarial syllabus. The subjects covered are the compilation of mortality and sickness rates, some discussion of the National Life Tables and elementary graduation such as the graphic and finite difference methods including, however, osculatory interpolation. More advanced topics such as selection, continuous exposed to risk formulae and multiple decrement tables are to be dealt with in a second volume. The bulk of the book is devoted to the calculation of the ordinary rate of mortality from life office data using the policy year, life year and calendar year methods. The treatment is both meticulous and exhaustive. The underlying notion used is that of risk time and the various formulae are built up by combining the appropriate portions of risk-time. This part of the book seems, however, to suffer from a complete lack of illustrative examples, which would probably drive the points home to the average student much more quickly. There is also a missed opportunity here in that the straightforward application of such methods to absenteeism rates, the repair rate of machines, breakdown of vehicles and so on is nowhere mentioned.

The later part of the book deals mainly with graduation. After the graphic method has been described some appropriate statistical tests for the observed data are given. For the χ^2 test the statement is made, p. 127, that 'the expected χ^2-total is, of course, the number of individual values of χ^2'. This remark seems, to a statistician, rather surprising unless arbitrary mortality rates are being used, which is not so in the case considered where they have been obtained by a graduation. The authors appear to realize that all is not well as they then go on to describe circumstances under which the degrees of freedom might be reduced by one.

As a whole the book suffers from the limitations imposed on it by being written for a specific syllabus. A more logical plan would require all the exposed to risk, including selection, to be grouped together. However, to an actuarial student the book will be a gold mine provided he concentrates on it carefully so as to extract the essence of the various procedures. At the same time it is a pity that the book was not made to appeal to a wider audience by the inclusion of a few non-actuarial examples. For this reason a more appropriate title to the book would be 'Mortality and Allied Investigations'.

P. G. MOORE

Numerical Methods. By ANDREW D. BOOTH. London: Butterworth's Scientific Publications. 1955. Pp. vii + 195. 35*s*.

The development of automatic digital computers has in recent years given remarkable impetus to the study of numerical analysis. This book provides a short, lucid survey of an extensive and rapidly growing field. Treated are: interpolation, numerical differentiation and integration, summation of series, ordinary and partial differential equations, simultaneous linear equations, non-linear algebraic equations, approximating functions, Fourier synthesis and analysis, integral equations. Monte Carlo methods for linear, differential and integral equations are briefly mentioned.

Orientation towards the modern machines is explicit, and an important feature is the revaluation of the older methods of hand calculation in the light of experience of automatic computation. The treatment is everywhere extremely brief. Even the largest chapter (38 pages) on simultaneous linear

equations provides a scanty treatment, with little special mention, for example, of simplified procedures for symmetric matrices. There is some account of the condition of matrices, but little general discussion of errors, more than ever important in lengthy machine computations. No exercises are provided.

Numerical analysis has not yet become a standard subject in undergraduate mathematical courses. It is therefore interesting to note that the present book is based on a course of lectures given by the author to final honours students at Birkbeck College. It should be helpful for teaching the basic principles; those who have a problem to solve and require a fuller treatment of some particular topic will find useful key references included in the text.

F. G. FOSTER

Numerical Analysis. By Z. KOPAL. London: Chapman and Hall Ltd. 1955. Pp. xiv + 556. 63*s*.

This book covers many aspects of numerical analysis and in particular such topics as interpolation, numerical differentiation and quadrature, differential equations and boundary value problems. The style of writing is clear and easy to follow yet no vigour is lost. The standard of mathematics assumed corresponds roughly with pure mathematics in the Advanced Level of the General Certificate of Education, and new concepts are fully explained as and when they occur. The author does not hesitate to use numerical examples to illustrate and drive home his various points if he feels that it is better than a mass of symbols by themselves, and this enhances the value of a book designed to illustrate numerical methods.

The proof of interpolation formulae utilizes algebraic methods following from the Lagrangian form of formulae where $n+1$ points of a function are used to fix a polynomial of order n. Alternative derivations of the main formulae using the quick and neat operational form are given in an appendix. At the end of each chapter there are some brief bibliographical notes which should enable anyone desiring to pursue the subject further to get a good start. There are also plenty of examples for the reader to work out, varying enormously in difficulty—some being straightforward applications of bookwork whilst others are more or less research problems. Answers are not given, which is a pity in the case of the more purely numerical problems.

Naturally any book of this nature has to be terminated at some point, and in this case the emphasis is on single variable techniques only. Thus in the very wide coverage given to differential equations the absence of any discussion, however brief, on partial equations is to be regretted. Perhaps also to complete the book a discussion of the inversion of matrices and an appendix with some tables of Besselian and Everett's interpolation coefficients might have been included. These, though, are minor criticisms of what is a very thorough and well-written text-book which should be of use to numerous applied scientists.

P. G. MOORE

Lectures on Functions of a Complex Variable. Edited by WILFRED KAPLAN. Michigan: University of Michigan Press. 1955. Pp. v + 433. $10.

This book is a symposium of thirty-one lectures on recent developments in complex variable theory—these having been given at the Conference held for this purpose at the University of Michigan in the summer of 1953. The lectures vary, from introductory outlines of different parts of the subject, to specialized contributions which are also to be published elsewhere in the appropriate pure- and applied-mathematical journals. A considerable proportion of the contributions are on 'topological analysis' and conformal mapping, but a variety of topics not coming under these heads are also dealt with. It is thus a book for the analyst who specializes in certain parts of his subject, and the statistical relevance is very indirect. Thus the contributions on 'The Distribution of Zeros of a Polynomial', 'Approximation by Polynomials', 'Expansion Theorems for Analytic Functions', etc., may well be of assistance in research into the mathematical functions of statistics but they could hardly be called statistically important.

More to the point perhaps, are the surprisingly large number of the papers which deal (in their applications at least) with harmonic and other potential functions. For in view of the mathematical equivalence of distributions of mass, charge, etc., with those of probability, it is remarkable that the vast corpus of potential theory has not borne more statistical fruit.

The book, however, lacks an index and it is very doubtful whether the publication of such a heterogeneous collection of essays (in quality and content) serves any useful purpose not already catered for by existing periodicals and journals. The price is prohibitive.

D. E. BARTON

OTHER BOOKS RECEIVED

1. **Introducción a los Metodos de la Estadística.** Segunda Parte. By SIXTO RIOS. Madrid: Instituto de Investigaciones Estadisticas del Consejo Superior de Investigaciones Científicas. 1954. Pp. 241.

2. **Anthropometry and Human Engineering,** from a Symposium held in The Netherlands in May 1954. London: Butterworth's Scientific Publications (on behalf of the Advisory Group for Aeronautical Research and Development, N.A.T.O.). 1955. Pp. 123. 21*s*.

3. **Mathematical Models of Human Behavior.** Proc. Symposium Dunlap and Associates Inc. Stanford, Connecticut: Dunlap and Associates Inc. 1955. Pp. vii + 103.

4. **Nonmetric Factor Analysis.** By C. H. COOMBS and R. C. KAO. University of Michigan Press. 1955. Pp. 63. $2.

5. **Graphical Method of Statistical Inference.** By M. MASUYAMA. Tokyo: Maruzen Co. Ltd. 1953. Pp. 83.

TRACTS FOR COMPUTERS

Department of Statistics, University College, London

I. Tables of the Digamma and Trigamma Functions. By ELEANOR PAIRMAN, M.A. Tables for summing $S = \sum_{i=1}^{\infty} \frac{1}{(p_1 i + q_1)(p_2 i + q_2) \cdots (p_n i + q_n)}$ where the p's and q's are numerical factors. *Price 5s. net.*

V. Table of Coefficients of Everett's Central-Difference Interpolation Formula. By A. J. THOMPSON, PH.D. *Second edition. Price 5s. net.*

VIII. Table of the Logarithms of the Complete Γ-Function (to ten decimal places) for Argument 2 to 1200 beyond Legendre's Range (Argument 1 to 2). By EGON S. PEARSON, D.SC. *Price 5s. net.*

IX. Log Γ (x) from $x = 1$ to 50·9 by intervals of 0·01. By JOHN BROWNLEE, M.D., D.SC. *Price 5s. net.*

X. On Quadrature and Cubature or on Methods of Determining Approximately Single and Double Integrals. By J. O. IRWIN, D.SC. *Price 7s. 6d. net.*

XII. Tables of the Probable Error of the Coefficient of Correlation. By KARL HOLZINGER, PH.D. *Price 5s. net.*

XIII. Bibliotheca Tabularum Mathematicarum, being a Descriptive Catalogue of Mathematical Tables. Part I. A, Logarithms of Numbers. By JAMES HENDERSON, PH.D. *Price 9s. net.*

XV. Random Sampling Numbers. By L. H. C. TIPPETT, M.SC., with a Foreword by KARL PEARSON. *Price 5s. net.*

XXIII. Tables of $\tan^{-1}x$ and $\log(1+x^2)$. To assist in the calculation of the ordinates of a Pearson Type IV curve. By L. J. COMRIE, PH.D. *Price 5s. net.*

XXIV. Random Sampling Numbers (2nd Series). By M. G. KENDALL and B. BABINGTON SMITH. *Price 5s. net.*

XXV. Random Normal Deviates. By HERMAN WOLD. *Price 5s. net.*

XXVI. Correlated Random Normal Deviates. By E. C. FIELLER, T. LEWIS and E. S. PEARSON. *Price 10s. 6d. net.*

Nos. II, III, IV, VI and VII are out of print

LOGARITHMETICA BRITANNICA

A standard Table of Logarithms to Twenty Decimal Places. By A. J. THOMPSON, Ph.D. (commenced in 1922 to commemorate the tercentenary of the publication of HENRY BRIGGS'S *Arithmetica Logarithmica*).

The nine separate sections of this Table have now been issued, and the complete work consisting of the logarithms of numbers 10,000–100,000, together with Dr Thompson's General Introduction (98 pp.) is now available in two bound volumes.

Price £8. 8s. 0d.

Issued by the CAMBRIDGE UNIVERSITY PRESS, Bentley House, LONDON, N.W.1
on behalf of the
DEPARTMENT OF STATISTICS, UNIVERSITY COLLEGE, LONDON
and obtainable from any bookseller

(*All Rights reserved*)

BIOMETRIKA. Vol. 43, Parts 1 and 2

CONTENTS

PAGE

Studies in the history of probability and statistics. II. The beginnings of a probability calculus. By M. G. KENDALL 1

On estimating the latent and infectious periods of measles. I. Families with two susceptibles only. By NORMAN T. J. BAILEY 15

The behaviour of an estimator for a simple birth and death process. By J. H. DARWIN . . . 23

Examination of a quantum hypothesis based on a single set of data. By S. R. BROADBENT . . 32

The number of new species, and the increase in population coverage, when a sample is increased. By I. J. GOOD and G. H. TOULMIN 45

A sequential test of randomness for events occurring in time or space. By D. J. BARTHOLOMEW . 64

On the moments of the maximum of partial sums of a finite number of independent normal variates. By A. A. ANIS 79

On the application to statistics of an elementary theorem in probability. By H. A. DAVID . . 85

χ^2 Probabilities for large numbers of degrees of freedom. By JOHN WISHART 92

The sampling distribution of a maximum-likelihood estimate. By J. B. S. HALDANE and SHEILA MAYNARD SMITH 96

Tests for randomness of points on a line. By D. E. BARTON and F. N. DAVID 104

Paired comparison designs for testing concordance between judges. By R. C. BOSE . . . 113

On the distribution of the largest or the smallest root of a matrix in multivariate analysis. By K. C. S. PILLAI 122

Tests of significance for the latent roots of covariance and correlation matrices. By D. N. LAWLEY 128

The variance of the mean of systematic samples. By R. M. WILLIAMS 137

Grouping methods in the fitting of polynomials to unequally spaced observations. By P. G. GUEST 149

On the joint distribution of the circular serial correlation coefficients. By G. S. WATSON . . 161

The approximate distribution of serial correlation coefficients. By H. E. DANIELS 169

Tests of hypotheses in the linear autoregressive model. II. Null distributions for higher order schemes: non-null distributions. By G. M. JENKINS 186

MISCELLANEA

A class of distributions for which the maximum-likelihood estimator is unbiased and of minimum variance for all sample sizes. By D. E. BARTON 200

Further critical values for the two-means problem. By W. H. TRICKETT, B. L. WELCH and G. S. JAMES 203

An approximation for the symmetric, quadrivariate normal integral. By J. A. MCFADDEN . . 206

Weighted probits allowing for a non-zero response in the controls. By M. J. R. HEALY . . 207

Treatment variances for experimental designs with serially correlated observations. By J. C. BUTCHER 208

The multivariate distribution of complex normal variables. By R. A. WOODING 212

Stationarity conditions for stochastic processes of the autoregressive and moving-average type. By J. WISE 215

Some properties of an angular transformation for the correlation coefficient. By B. I. HARLEY . 219

Note on the moment-problem for unimodal distributions when one or both terminals are known. By C. L. MALLOWS 224

On inverting a class of patterned matrices. By S. N. ROY and A. E. SARHAN 227

A note on the risks of error involved in the sequential ratio test. By J. MEDHI 231

CORRIGENDA 235

REVIEWS

LANCELOT HOGBEN'S 'Choice and Chance by Cardpack and Chessboard', vol. II 236

R. M. THRALL, C. H. COOMBS and R. L. DAVIS (edd.), 'Decision Processes' 236

R. R. BUSH and F. MOSTELLER'S 'Stochastic Models for Learning' 237

MORTIMER SPIEGELMAN'S 'Introduction to Demography' 237

H. W. HAYCOCKS and W. PERKS'S 'Mortality and Other Investigations', vol. I 238

ANDREW D. BOOTH'S 'Numerical Methods' 238

Z. KOPAL'S 'Numerical Analysis' 239

WILFRED KAPLAN (ed.), 'Lectures on Functions of a Complex Variable' 239

OTHER BOOKS RECEIVED 240

Printed in Great Britain at the University Press, Cambridge (Brooke Crutchley, University Printer)

PAGE
1
15
23
32
45
64
79
85
92
96
104
113
122
128
137
149
161
169
186
200
203
206
207
208
212
215
219
224
227
231
235
236
236
237
237
238
238
239
239
240